Approches macro-écologique et phylogéographique pour démêler facteurs et processus responsables des patrons de biodiversité aquatique souterraine en Europe

David Eme

N° d'ordre : 134 – 2104 Année 2014

THESE

présentée devant

L'UNIVERSITE CLAUDE BERNARD LYON 1

pour l'obtention du

DIPLOME DE DOCTORAT

(arrêté du 7 août 2006)

ECOLE DOCTORALE 341 - E2M2

Évolution Écosystèmes Microbiologie Modélisation

Présentée et soutenue publiquement le 16 Juillet 2014

M. David EME

TITRE :

Approches macro-écologique et phylogéographique pour démêler facteurs et processus responsables des patrons de biodiversité aquatique souterraine en Europe.

Directeur de thèse : Christophe DOUADY

JURY M. Sylvain DOLEDEC, examinateur
M. Christophe DOUADY, examinateur et encadrant
M. Bernard HUGUENY, rapporteur
M. Fabien LEPRIEUR, examinateur
M. Florian MALARD, examinateur et encadrant
M. Ludovic ORLANDO, rapporteur
M. François POMPANON, examinateur

Résumé Un ensemble de disciplines tente de comprendre les causes de la distribution de la biodiversité à la surface de la terre. Cette thèse, à l'interface entre macro-écologie et phylogéographie, démêle le rôle relatif des différents facteurs environnementaux et des processus contrôlant la diversité des crustacés aquatiques souterrains en Europe. L'utilisation d'un modèle biologique souterrain permet d'écarter l'effet de la saisonnalité thermique, omniprésente dans les milieux de surface. L'action de multiples facteurs – plus particulièrement la disponibilité des ressources trophiques et l'hétérogénéité environnementale – et les variations régionales de leur importance relative fournissent l'explication la plus parcimonieuse des patrons de richesse. Ce résultat s'oppose au paradigme du rôle prépondérant du processus d'extinction causé par les fortes oscillations climatiques du Pléistocène en Europe du nord. Toutefois, ces oscillations ont très probablement sélectionné des organismes mobiles qui participent à l'augmentation de la taille moyenne des aires de répartition des espèces avec la latitude. La reconstruction de la dynamique des aires de distribution montre que la dispersion est un processus très hétérogène entre et au sein des espèces. Elle interviendrait lors de courtes fenêtres temporelles entre lesquelles l'adaptation locale tendrait au contraire, à contrecarrer les capacités de dispersion. Enfin, ce travail propose des pistes de réflexion afin d'expliquer plus précisément, à partir de données moléculaires supplémentaires et d'outils génomiques, les variations géographiques des taux de diversification et de substitution à l'échelle continentale.

Mots clés Patrons de biodiversité, macro-écologie, phylogéographie, crustacés, eau souterraine, carte d'habitat souterrain, reconstruction d'aire ancestrale, dispersion, adaptation locale, multi-causalité, non-stationnarité, richesse spécifique, taille des aires de distribution, beta diversité, niche réalisée/fondamentale, diversité cachée, *Proasellus*, Europe.

Title Methods in macroecology and phylogeography for disentangling factors and processes shaping groundwater biodiversity patterns in Europe.

Abstract A set of disciplines attempt to understand causes of biodiversity patterns on the earth. This thesis, at the frontier between macroecology and phylogeography, disentangles the relative influence of environmental factors and processes shaping groundwater crustacean diversity in Europe. Groundwater habitats offer useful case studies for avoiding the effect of thermal seasonality, which is pervasive in surface ecosystems. The influence of multiple factors – especially productive energy and spatial heterogeneity – and regional variation in their relative importance provide the most parsimonious explanation of species richness patterns. This result undermines the prominent role attributed to the disproportionate extinction of species in northern European regions with high historical climate oscillations. However, these oscillations have probably selected vagile species which contribute to the increase in median range size of species with latitude. Reconstructing range dynamics shows that dispersal is a heterogeneous process within and among species. It may occur during short time windows between which local adaptation favors specialization. Finally, I suggest several research avenues using molecular data and genomic tools for understanding geographical variation in diversification and substitution rates at continental scale.

Keywords: Biodiversity patterns, macroecology, phylogeography, crustaceans, groundwater, groundwater habitat map, ancestral range reconstruction, dispersal, local adaptation, multi-causality, non-stationnarity, richness, range size, beta diversity, realized/fundamental niches, cryptic diversity, *Proasellus*, Europe.

Université Claude Bernard - Lyon I, CNRS, UMR 5023 - LEHNA (Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés), 43 Boulevard du 11 novembre 1918, Bâtiments Darwin C et Forel, 69622 Villeurbanne Cedex

Remerciements

Les remerciements qui vont suivre sont très simples, épurés et cachent une certaine pudeur dans l'expression de mes émotions. Autrement dit, toute ma reconnaissance que j'ai pour toutes les personnes qui m'ont accompagné de près ou de loin au cours de ce travail de thèse ne s'exprimera malheureusement pas à sa juste valeur dans les lignes qui vont suivre. Je m'en excuse par avance, la vérité est ailleurs !

Je tiens à remercier mes encadrants, Christophe Douady et Florian Malard pour leurs investissements énorme au quotidien et l'ensemble de leurs apports scientifiques et humains sur ce travail. Ces apports sont considérables et il m'est difficile de les verbaliser… Je crois qu'ils ont, « simplement », participés à la construction de ma colonne vertébrale de jeune scientifique, UN GRAND MERCI à vous !

Je tiens à remercier très sincèrement les membres du jury de thèse qui ont accepté avec beaucoup de gentillesse d'évaluer mon travail. Merci à Ludovic Orlando et Bernard Hugueny pour vos retours très positifs sur le manuscrit, et merci à tous les examinateurs Fabien Leprieur, François Pompanon et Sylvain Dolédec, pour cet après-midi de science au cours duquel j'ai pris beaucoup de plaisir à échanger avec vous.

Je remercie très chaleureusement :

Lara Konecny pour son aide immense, sa méticulosité et sa rigueur dans l'acquisition des séquences ADN. Merci à Pauline Jean et Céline Colson-Proch pour avoir été à l'origine de nombreuses données et analyses pour le papier « valdensis ».

Toutes les personnes de l'équipe E3S et du LEHNA actuelles ou passées que j'ai côtoyé depuis un certain temps maintenant.

Toutes les personnes qui m'ont aidé professionnellement au cours de cette thèse : Sanja Gottstein, Maja Zagmajster, Cene Fiser, Teo Delic, Pierre Marmonier, Christian Feld, Florian Pletterbauer, Thibault Datry, Thierry Oberdorff, Clément Tisseuil, Jean François Cornu, David Culver, Chris Robinson, Diana Galassi, Fabio Stoch, Pablo Tedesco, Julien Clavel, Gilles Escarguel, Emmanuel Paradis, Mike Monaghan, Sébastien Brosse, Nicolas Bailly, Tristan Lefébure, Dominique Martin, Nadjette Houriez…et j'en oublie.

Amis et camarades thésards, PostDocs, stagiaires, masters croisés au LEHNA et ailleurs, Jules SB, Natacha LP, Capucine SB, Bertrand B, Stéphanie P, Julien B, Julien G, Virginie «Kath» DS, Alexandre T, Pierre C, Anne N, Claire M, Cécile C, Jonathan C, Evelyne T, Ingrid B, Julien «corsinou» V, Lucile M, Morgan B, François K, Clémentine F, Natacha F et Erik, Marie Laure C, Jérémy L, Guillaume F, Christophe P, Alice M, Jérôme P, Loïc B, Thomas W, Mathieu T, Monica CCL, Javier CL, Yvo, Mathieu P, Raoul TK, Nathanaëlle S, Julie T, Antonin V, Ludovic G, Charlotte G, Barbara LR, Melissa DW, Simon N, Florent A, Arnaud F, Emilien L, Loïc T, Jehanne OC, Julia D, Hugo C, Jérôme G, Silvia VM, Felix PM, Jérémy V, Maxence F, Samuel S, Nicolas P, Delphine C, Chafik M….et toutes les personnes oubliées dans ces quelques lignes.

Ma famille et ma belle-famille qui ont été admirable pour faire face à mes absences intellectuelles !

Et bien sûr Chloé.

Table des matières

Introduction

Comprendre le rôle des facteurs environnementaux et des processus impliqués dans la répartition de la biodiversité à la surface du globe est une thématique clé, développée par la macro-écologie et la phylogéographie, champs disciplinaires à l'interface entre écologie et évolution (Brown, 1995 ; Gaston, 2000 ; Jetz *et al.* 2012). La macro-écologie correspond à l'analyse des facteurs et des processus responsables de la distribution de multiples espèces sur de larges échelles de temps et d'espace (Brown, 1995). La phylogéographie étudie les principes et les processus gouvernant la distribution géographique des lignées généalogiques au sein des espèces (Avise, 2000). Le champ d'action de ces disciplines couvre un vaste programme de recherche, initié de longue date par Darwin et Wallace, qu'il est plus que jamais pertinent d'explorer dans un contexte de dérèglement climatique mondial (Gaston, 2000) Par ailleurs, les demandes sociétales de mise en place de mesures de protection de la biodiversité à large échelle spatiale sont de plus en plus pressantes (Convention sur la Diversité Biologique, CBD, 2010). Ainsi, le troisième objectif stratégique fixé à l'horizon 2020 par la conférence internationale du Programme des Nations Unies sur la biodiversité de 2010 (Aichi, Japon), stipule « d'améliorer l'état de la diversité biologique en sauvegardant les écosystèmes, les espèces et la diversité génétique » (CDB, 2010, http://www.cbd.int/sp/targets/). Toutefois, proposer des plans de gestion cohérents permettant de lutter contre l'érosion de la biodiversité ne peut être réalisé sans une connaissance approfondie des facteurs environnementaux et des processus impliqués (Boakes *et al.* 2010 ; Jetz *et al.* 2012). Cette connaissance est indispensable pour modéliser et prédire les changements de structure, de répartition et de fonctionnement de la biodiversité en réponse à l'intensification des facteurs de forçage liés au changement global.

La compréhension des liens de causalité entre facteurs environnementaux, processus et patrons de biodiversité est une tâche particulièrement ardue dans la mesure où un même patron peut résulter de l'action de différents processus, eux-mêmes sous l'influence de différents facteurs (Lawton, 1999 ; Gaston, 2000 ; Vellend, 2010). En plus de cette difficulté inhérente à la problématique, cette tâche est actuellement confrontée à deux obstacles. Premièrement, l'étendue des questionnements a entrainé un cloisonnement des champs disciplinaires, qu'il convient aujourd'hui de mettre en interaction, voire de réunifier (Graham & Fine, 2008 ; Cavender-Bares *et al.* 2009 ; Nogues-Bravo 2009 ; Hickerson *et al.* 2010 ; Lavergne *et al.* 2010 ; Jenkins & Ricklefs, 2011 ; Ricklefs & Jenkins, 2011 ; Mouquet *et al.* 2012 ; Chave, 2013 ; Marske *et al.* 2013 ; Thuiller et al. 2013). Bien qu'en pleine mutation, l'écologie s'est longtemps fondée sur une approche hypothético-déductive expérimentale et réductionniste, contrairement à l'évolution qui a d'avantage suivi une approche inductive holiste et non expérimentale (Brown, 1995). De par leurs contraintes propres, ces approches se sont souvent intéressées à des échelles spatiales et temporelles distinctes mettant en confrontation des processus et des facteurs qui pouvaient pourtant agir de concert (Shmida & Wilson, 1985). Deuxièmement, les connaissances actuelles portant sur les liens entre facteurs environnementaux et processus proviennent d'un nombre relativement restreint de taxons (mammifères, oiseaux, amphibiens …), majoritairement issus de milieux terrestres et marins (Beck *et al.* 2012). Ces taxons ont été utilisés essentiellement en raison de la disponibilité des jeux de données et non en fonction de la pertinence des hypothèses à tester. Par exemple, les études menées sur les milieux de surface sont toutes confrontées à la difficulté d'estimer l'échelle temporelle à laquelle la variabilité climatique est la plus susceptible de façonner les patrons de biodiversité (Morueta-Holmes *et al.* 2013 ; Veter *et al.* 2013). Les réponses à de nombreuses questions restées en suspens pourraient provenir de l'élargissement des recherches à des taxons occupant des milieux aux caractéristiques environnementales singulières mais particulièrement pertinentes (Gaston *et al.* 1998).

La genèse de cette thèse émane de ce double constat et de la question qui en découle : l'étude à la frontière entre deux disciplines d'un modèle biologique original ne peut-elle pas apporter un éclairage supplémentaire à la compréhension des liens entre facteurs, processus et patrons de biodiversité ? Les écosystèmes aquatiques souterrains continentaux ne sont certes pas des macrocosmes naturels mais la réduction des sources de variabilité climatique et écologique (stabilité thermique au cours de la saison, réseaux trophiques et communautés simplifiées…) est une caractéristique particulièrement attrayante lorsqu'il s'agit de démêler les liens entre facteurs environnementaux et processus. Alors que les eaux continentales souterraines représentent près de 97% des eaux continentales libres (Castany, 1998), l'essentiel de nos connaissances sur les déterminants de la biodiversité des milieux aquatiques d'eau douce est fondé quasi exclusivement sur les organismes de surface. L'importance des eaux souterraines dans le fonctionnement des écosystèmes terrestres est considérable, notamment pour l'homme comme source d'eau potable et d'irrigation pour l'agriculture (Morris *et al.* 2003). Pourtant, la diversité biologique de ce milieu reste largement inexplorée, tout comme les facteurs et les processus qui régissent la répartition de sa diversité (Gibert *et al.* 2009 ; Deharveng *et al.* 2009).

L'objectif de cette thèse est précisément d'identifier et de quantifier l'influence relative des facteurs environnementaux et des processus impliqués dans la distribution spatiale de la biodiversité des eaux souterraines continentales européennes. Ce travail de recherche intègre plusieurs champs disciplinaires. Il comporte des études de macro-écologie visant à inventorier et à estimer l'influence relative de multiples facteurs environnementaux sur la diversité spécifique des crustacés aquatiques souterrains à l'échelle européenne. Il comporte également des études menées à des niveaux d'organisation biologique inférieurs (espèces morphologiques) qui couplent des approches phylogéographique, physiologique et écologique (modélisation de niche) afin d'évaluer plus précisément le rôle des processus de dispersion et de sélection.

Architecture du manuscrit

Le premier chapitre de ce manuscrit présente le cadre conceptuel et les objectifs de ce travail de thèse. Il comporte quatre parties. La première clarifie certains termes, décortique les patrons, les processus, et les facteurs ainsi que les liens qu'ils entretiennent. La seconde partie examine comment ces liens sont abordés par différents champs disciplinaires. La troisième expose l'état des connaissances sur les processus et les facteurs environnementaux gouvernant la macro-distribution de la biodiversité aquatique souterraine, lorsque ce travail de recherche a débuté. Enfin, les objectifs et les hypothèses de cette thèse sont présentés dans la quatrième partie.

Le second chapitre analyse l'influence relative des différents facteurs environnementaux (histoire, énergie actuelle et hétérogénéité spatiale) sur la richesse spécifique et la taille des aires de répartition des crustacés aquatiques souterrains en Europe. Ce chapitre comporte trois parties. La première expose les résultats d'un travail aboutissant à la réalisation de la première carte typologique des habitats aquatiques souterrains en Europe. Celle-ci permet de calculer des métriques de disponibilité et d'hétérogénéité d'habitats propres à ce milieu, qui sont indispensables aux analyses développées lors des deux parties suivantes. La deuxième partie s'appuie sur un jeu de données spécifiquement assemblé dans le cadre de cette thèse regroupant l'ensemble des occurrences de crustacés aquatiques souterrains en Europe. Ce jeu de données issu d'un travail collaboratif est dans un premier temps utilisé pour documenter les patrons géographiques de richesse spécifique, la taille des

aires de répartition des espèces et de béta diversité (dissimilarité de composition en espèces entre sites, cf. chapitre 1). Une analyse permet ensuite de tester l'influence relative de l'hétérogénéité spatiale et de la variabilité climatique à deux échelles distinctes de temps sur le patron de taille des aires de répartition des espèces en Europe. Dans la troisième partie, ce même jeu de données distributionnelles est utilisé pour tester l'influence relative de la variabilité climatique à long terme, de la disponibilité en énergie et de l'hétérogénéité spatiale sur le patron de richesse spécifique. L'analyse effectuée permet par ailleurs de tester comment l'influence relative de ces trois facteurs varie régionalement.

Le troisième chapitre évalue le rôle de la dispersion et des colonisations postglaciaires chez les organismes souterrains. Il utilise des outils moléculaires et une approche de phylogéographie comparative pour tester l'hétérogénéité de la dispersion entre et au sein de plusieurs espèces d'isopodes et reconstruire la dynamique temporelle des aires de répartition de trois d'entre elles.

Le quatrième chapitre croise des approches de phylogéographie, de modélisation de niche et d'expérimentation physiologique afin d'évaluer le rôle relatif des processus de dispersion et de sélection et de leur interaction dans la mise en place de l'aire de répartition d'une espèce d'isopode.

Le cinquième chapitre synthétise et discute les résultats marquants de cette thèse tant dans le domaine de l'écologie en général que dans celui de la biologie souterraine et développe les perspectives qui en découlent. La robustesse des patrons documentés ainsi que deux résultats de ce travail portant sur l'importance de l'énergie à large échelle spatiale et le rôle de la dispersion en milieu souterrain, y sont l'objet d'un examen critique. Enfin, trois perspectives sont développées sous la forme d'un programme de recherche visant à approfondir la compréhension des processus et des liens qu'ils entretiennent avec les facteurs environnementaux, tout en conservant la généralité offerte par le cadre macro-écologique.

Les annexes présentent le matériel supplémentaire des cinq articles inclus dans ce travail de thèse. Elles comprennent également un article auquel j'ai participé en tant que co-auteur au sein d'un groupe de travail du programme européen BioFresh à l'origine du financement de ma thèse. Cet article soumis à la revue *Freshwater Biology,* a évalué dans un cadre multi-groupes (« poissons », mollusques, macrophytes, invertébrés benthiques, crustacés souterrains) et multi-écosystèmes (lotique, lentique et souterrain) l'importance relative des pressions anthropiques exercées par l'utilisation du paysage et des facteurs naturels sur les patrons de biodiversité des eaux douces en Europe.

Programme de recherche soutenant ce travail

Ce travail de recherche s'inscrit dans le cadre du programme européen BioFresh (7th EU Framework Programme, Contract N° 226874, coordonnateur : K. Tockner) dont l'objectif est de centraliser et de rendre disponible l'ensemble des bases de données décrivant la distribution, le statut et les tendances de la biodiversité des eaux douces dans le monde. Cet objectif se décline comme suit :

- fournir une plateforme permettant de visualiser dans l'espace les informations sur le statut et les tendances de la biodiversité et les écosystèmes d'eaux douces.
- accroitre la sensibilisation sur l'importance de la biodiversité des eaux douces et les services écosystémiques fournis par les milieux d'eau douce.

- prédire les réponses futures de la biodiversité face aux multiples pressions exercées suite au changement global.

Pour répondre à ces objectifs, le projet a été divisé en 8 tâches distinctes réparties parmi 19 institutions partenaires. Mon travail de thèse s'insère plus particulièrement au sein de deux tâches qui avaient pour but de documenter et de comprendre les patrons actuels et passés de la biodiversité des eaux douces.

Dans le cadre de ce programme, cette thèse avait pour ambition : i) d'utiliser un écosystème disposant de caractéristiques particulières pour nous aider à comprendre les liens entre patrons, facteurs et processus ; ii) de combler le retard de connaissances sur les patrons et les déterminants de la biodiversité des eaux douces souterraines continentales par rapport aux écosystèmes de surface. Il s'agissait plus particulièrement de fournir des cartes de distribution de la biodiversité aquatique souterraine en Europe et de proposer des hypothèses crédibles soutenant ces patrons de diversité.

Les missions de terrain et l'acquisition des données génétiques utilisées dans cette thèse ont été financées par l'Agence Nationale de la Recherche (ANR 08JCJC012001, « DEEP », coordonnateur C. Douady). Le but de ce programme était de démêler les processus évolutifs et écologiques responsables des patrons de biodiversité aquatique souterraine. Enfin, le partenariat Hubert Curien (HPC) Franco Slovène Proteus (No. 31199UM) a permis de financer en partie un travail collaboratif avec des collègues slovènes de l'université de Ljubljana ayant abouti aux articles 2 et 3. Ce programme a pour objectifs de documenter les processus et les facteurs environnementaux impliqués dans les patrons de diversité et d'évaluer les changements de patrons de diversité en passant d'une identification morphologique à une identification moléculaire de la diversité.

Chapitre1) Cadre conceptuel et objectifs

1.1) Liens entre patrons de biodiversité, processus et facteurs environnementaux

1.1.1) Quelques éléments de clarification

Les liens de causalité entre facteurs, processus et patrons de biodiversité constituent de véritables boîtes noires, car un même patron de biodiversité peut être issu de différents processus, eux-mêmes sous l'influence de différents facteurs (Lawton, 1999 ; Gaston, 2000 ; Vellend, 2010). Alors que les processus représentent bien souvent les objets d'intérêts, leurs inférences restent le plus souvent largement indirectes *via* des approches corrélatives entre facteurs environnementaux et les nombreuses métriques quantifiant la biodiversité. Toutefois, avant de rentrer dans l'étude de cette complexité, il me semble nécessaire d'expliciter l'abondante terminologie relative à la description des déterminants de la biodiversité. La clarification des termes employés au cours de ce manuscrit relève davantage de ma représentation et de mon cheminement personnels plutôt que de définitions consensuelles.

La biodiversité exprime la diversité des entités du vivant à différents niveaux d'organisation biologique allant du gène, aux écosystèmes en passant par les populations (ensemble des individus d'une espèce vivant en un lieu et un temps donné, Waples & Gaggiotti, 2006), l'espèce et la communauté (un ensemble d'organismes appartenant à plusieurs espèces vivant en un lieu et un temps donné, Vellend, 2010) (Fig. 1). Cette biodiversité est quantifiée à chacun des niveaux d'organisation biologique à partir de métriques (e.g. richesse allélique, richesse spécifique), dont on verra par la suite que le choix est loin d'être neutre. La représentation dans l'espace et/ou le temps des variations de ces métriques constituent des patrons de biodiversité. Ces patrons reflètent l'effet des facteurs environnementaux sur la contribution relative de différents processus à la distribution de la biodiversité. Par exemple, de faibles ressources trophiques (facteur) limitent la taille des populations, favorisant ainsi le rôle de la dérive par rapport à celui de la sélection (processus) dont les conséquences peuvent se traduire par une perte de diversité génétique (patron). Inversement, la biodiversité agit sur l'environnement et peut de ce fait moduler l'effet des processus. Ces phénomènes de rétroaction ne seront pas abordés dans ce travail.

Processus et facteurs peuvent chacun être représentés sous la forme d'une hiérarchie (Fig.1). Au niveau basal de la hiérarchie des processus, quatre forces évolutives (mutation, migration, sélection et dérive) façonnent la diversité biologique (théorie neutraliste de l'évolution moléculaire et de la génétique des populations, Kimura, 1983). Au niveau hiérarchique supérieur, celui des communautés, Brown (1995) précise que trois processus façonnent le nombre d'espèces en un lieu ou un temps donné : la spéciation (création), l'extinction (disparition) et la dispersion (déplacement). Vellend (2010) a clarifié l'écologie des communautés en la présentant dans un cadre régi par quatre processus : la spéciation, la sélection, la dérive et la dispersion (Vellend, 2010). Notons à cet égard que les deux approches ne sont pas incompatibles ; il s'agit uniquement de deux perceptions différentes. Pour Vellend (2010), l'extinction est uniquement le fruit d'un jeu entre sélection et dérive. Plus généralement, la spéciation et l'extinction peuvent être perçues comme la résultante des quatre forces évolutives au niveau d'organisation biologique inférieur. Dans ces conditions, les facteurs environnementaux agissent sur les processus du niveau inférieur et les processus aux niveaux hiérarchiques supérieurs découlent directement de ces actions (Avise *et al.* 1987). Dans le cadre de la macro-évolution, certains auteurs, défendent l'idée que les forces évolutives agissant au niveau de

l'individu (micro-évolution) ne sont pas les seules à participer à la genèse des patrons de biodiversité à large échelle spatiale et temporelle (Gould & Lewontin 1979 ; Erwin, 2000). Bien que cette vision soit âprement critiquée, la sélection pourrait agir sur d'autres entités que l'individu, par exemple la sélection de groupes agirait chez les organismes eu-sociaux (Wilson & Hölldobler, 2005). Entrer dans ce débat entre processus micro- et macro-évolutif ne m'apparait pas nécessaire dans le cadre de cette thèse.

Par commodité, il est également possible de hiérarchiser en quatre classes de facteurs environnementaux (Fig. 1). Par exemple, l'énergie représente un facteur intégratif qui regroupe l'énergie ambiante et l'énergie productive (Evans *et al.* 2005, cf. ci-après pour une définition de ces composantes de l'énergie). De la même façon, l'hétérogénéité environnementale représente un facteur intégratif qui peut se décliner sous la forme d'une hétérogénéité topographique et d'une diversité d'habitats (Kerr *et al.* 2001).

Les hiérarchies de processus et de facteurs proposées dans ce paragraphe ne s'appuient pas sur des échelles spatiale ou temporelle afin d'éviter le jeu délicat des correspondances entre les hiérarchies d'échelles spatiale et temporelle et les hiérarchies de facteurs et de processus. Nous ne rentrerons pas dans cette problématique au cours de ce manuscrit dans la mesure où les questionnements portent sur deux niveaux d'organisation biologique, la population et la communauté, abordés uniquement à large échelle spatiale (de la région au continent). Il est nécessaire de préciser que la notion d'échelle spatiale revêt deux composantes. La résolution définit l'unité spatiale élémentaire d'investigation aussi appelé grain spatial, et l'étendue correspond à la zone géographique étudiée (Rahbek, 2005).

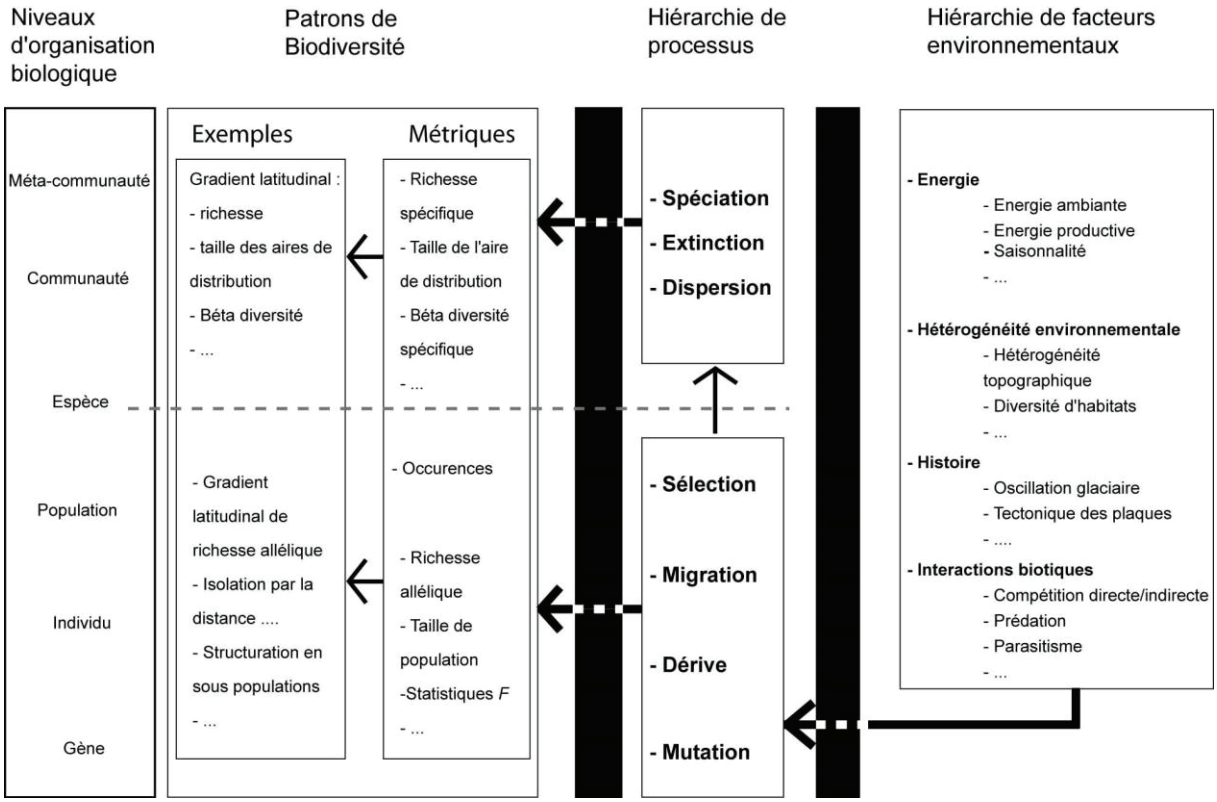


Figure 1 : Schéma conceptuel des liens entre patrons de biodiversité et hiérarchies de processus et de facteurs environnementaux.

1.1.2) Tant de taxons et si peu de patrons à large échelle spatiale

Quelques patrons récurrents

Malgré le très grand nombre de groupes taxonomiques et leur grande diversité morphologique, fonctionnelle ou comportementale, la majorité d'entre eux expriment à de larges échelles spatiales des distributions géographiques étonnamment similaires. La congruence des patrons de distribution entre taxons à longtemps laissé penser qu'un mécanisme universel «The Holy Grail of Ecology », en serait la cause (Davies *et al.* 2011). Parmi ces quelques patrons récurrents, citons le gradient latitudinal de richesse connu de très longue date (Darwin 1862, Wallace 1878), ou celui d'augmentation de la taille des aires de répartition avec la latitude connu initialement sous le nom de règle de Rapoport (Stevens, 1989) ou « effet de Rapoport » (Rohde, 1996).

Le gradient latitudinal de richesse a été mis en évidence chez une très grande majorité de groupes animaux et végétaux avec quelques exceptions, tels les trématodes parasites de « poisson » (Poulin, 1995), les oiseaux et mammifères marins (oiseaux pélagiques et mammifères pinnipèdes), et les plantes angiospermes aquatiques (Willig *et al.* 2003). La richesse spécifique est maximale autour de l'équateur dans la zone intertropicale et elle décroit en allant vers les pôles (Willig *et al.* 2003 et voir Fig. 2). , Ce gradient en place depuis au moins 70 millions d'années a suscité et suscite encore un vif intérêt pour la communauté scientifique. Toutefois, la quête des mécanismes à l'origine de cette relation a fait émerger une quantité impressionnante d'hypothèses vraisemblables pour l'expliquer (plus de 30), plutôt que de valider un mécanisme universel (voir review de Willig *et al.* 2003 ; Hillebrand, 2004 ; Mittellbach *et al.* 2007).

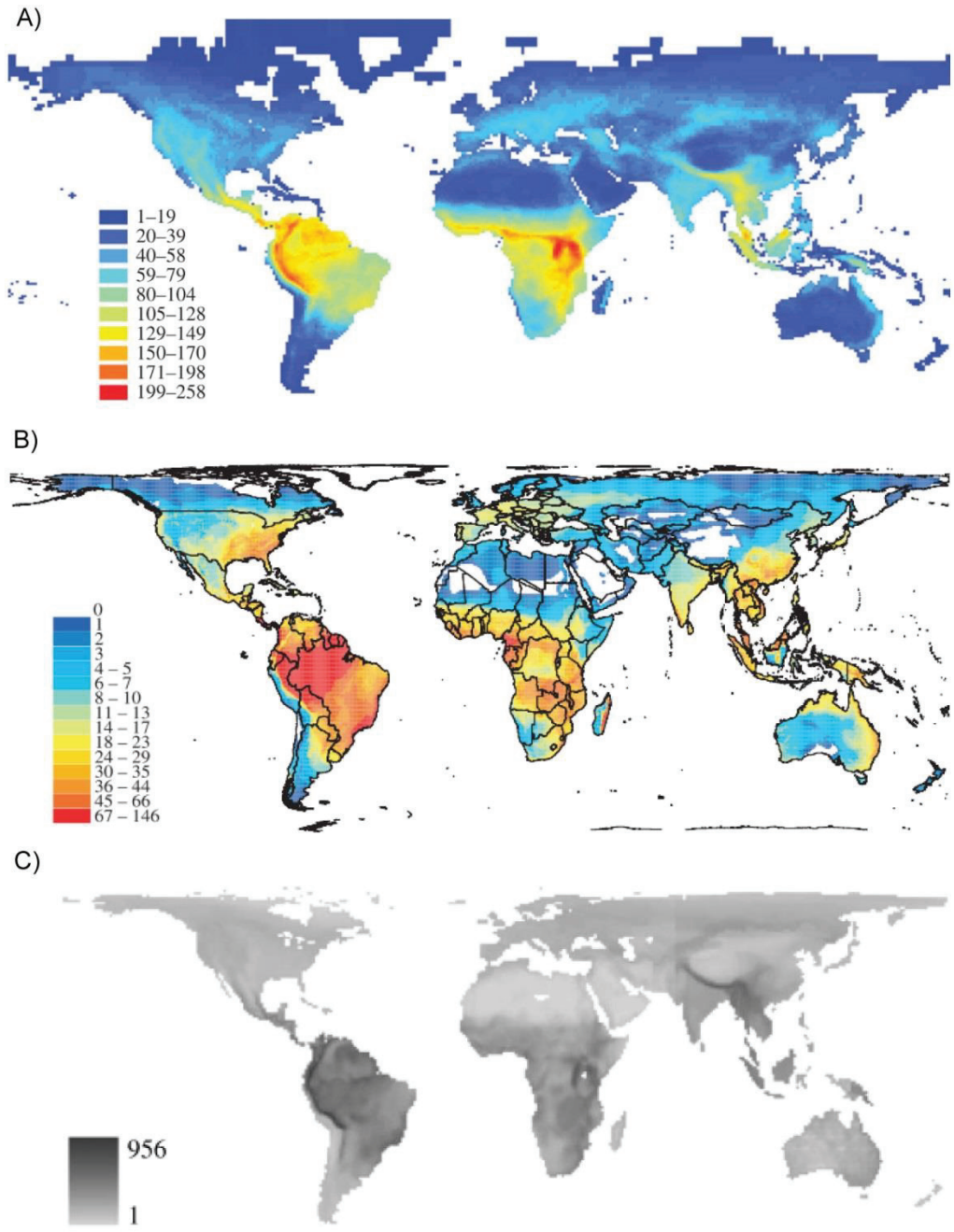


Figure 2 : Patrons globaux de la richesse spécifique actuelle **A)** des mammifères d'après Davies & Buckley (2011), **B)** des amphibiens, d'après Buckley & Jetz (2007), **C)** de l'avifaune d'après Davies et al. (2007a). Les patrons de richesse spécifique de ces trois groupes sont congruents à l'échelle globale.

La règle de Rapoport – l'augmentation de la taille moyenne des aires de répartition des espèces avec la latitude – représente un autre patron particulièrement célèbre (Stevens, 1989 ; Gaston *et al.* 1998 ; Veter *et al.* 2013). En revanche, contrairement au patron latitudinal de richesse, sa généralité a été source de controverses (Erwin 1989 ; Rohde, 1996). Rohde (1996) suggère que cette augmentation latitudinale des tailles des aires de répartition ne peut pas expliquer le gradient latitudinal de richesse, car il ne s'agit que d'un patron « local » limité au Néarctique (Amérique du Nord) et au Paléarctique (Eurasie) à des latitudes supérieures à 40°N. Dès lors, cette règle a été déclassée au rang d'effet par certains auteurs (Rohde, 1996).

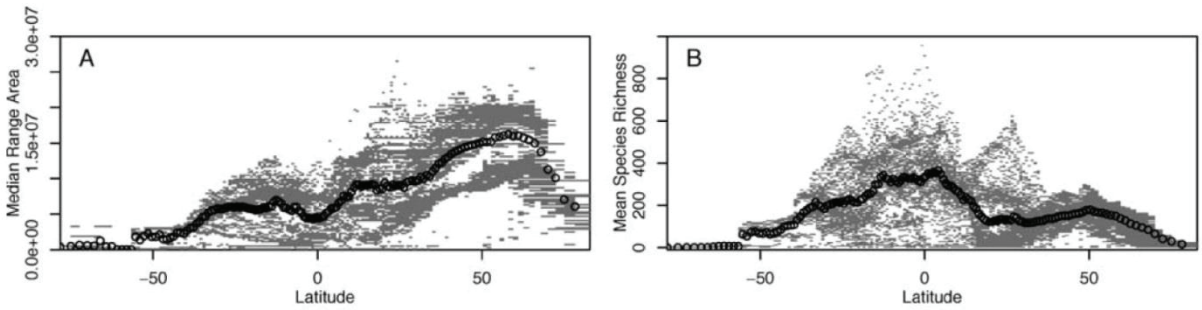


Figure 3 : Patrons latitudinaux pour l'avifaune à l'échelle globale **A)** de la taille médiane des aires de répartition supportant la règle de Rapoport dans l'hémisphère nord, **B)** de la richesse spécifique,. D'après Orme *et al.* (2006).

Notons que la baisse de la diversité avec la latitude a également été documentée pour la diversité génétique (haplotypique) populationnelle. Pour un ensemble d'espèces de plantes, de vertébrés et d'invertébrés dits méditerranéens, les populations des péninsules sud européennes disposent d'une plus grande variabilité génétique que celles localisées plus au nord (Hewitt, 1996, 1999).

Le choix des métriques est primordial

Les patrons sont construits à partir de métriques quantifiant la biodiversité qui disposent de propriétés différentes. Le choix de ces métriques en fonction des processus d'intérêts constitue une étape déterminante. Je ne présenterai qu'un nombre limité de métriques d'incidences (présence/absence) utilisées au cours de ce travail. Ainsi, les indices de diversité (Shannon ; Simpson ; entropie quadratique…) prenant en compte l'abondance des espèces ou les indices de diversité fonctionnelle ne seront pas abordés (Mouchet *et al.* 2010 ; Schleuter *et al.* 2010).

Les métriques quantifiant la biodiversité ont été développées dans le cadre d'une décomposition hiérarchique multiplicative ou additive de trois composantes de la diversité ; alpha, béta, et gamma (Jost 2007 ; Baselga, 2010a). Les composantes alpha et gamma sont quantifiées par la richesse qui énumère les entités biologiques appartenant à un même niveau d'organisation biologique (allèles, espèces, communautés…) dans une surface et en un temps donné. Les diversités alpha et gamma expriment la richesse respectivement aux échelles locale et régionale (richesse de l'ensemble des entités locales). La richesse est la résultante du chevauchement des aires de répartition des entités biologiques (Borregaard & Rahbek, 2010). Elle dispose de propriétés intégratives permettant d'approcher des processus de spéciation, d'extinction et dans une certaine mesure de dispersion. En revanche, elle présente l'inconvénient de perdre l'identité des entités biologiques.

La diversité béta représente les différences de composition entre assemblages locaux d'entités biologiques. Déclinée à travers de nombreux indices, elle offre les avantages de conserver l'identité des entités biologiques et, en fonction des indices considérés, d'être indépendante de la richesse (Koleff *et al.* 2003 ; Tuomisto, 2010a,b ; Anderson *et al.* 2011). Le lien qu'elle établit entre richesse locale et régionale rend cette métrique particulièrement utile pour inférer des processus de dispersion. Le développement récent des d'indices décomposant la béta diversité en deux composantes additives, le renouvellement spatial (« spatial turnover») et la diversité emboitée (« nesteness resultant dissimilarity », ou « nestedness » sur la Fig. 4) offre de nouvelles perspectives de recherche (voir la Figure 4, Baselga, 2010b, 2012; Carvalho *et al.* 2011; Leprieur *et al.* 2011 ; Hortal *et al.* 2011 ;

Svenning *et al.* 2011). Le renouvellement spatial permet d'inférer des processus de spéciation, alors que la composante liée à la diversité emboitée permet de documenter des gradients de contraintes à la dispersion ou des gradients d'extinction différentielle (Baselga, 2010b).

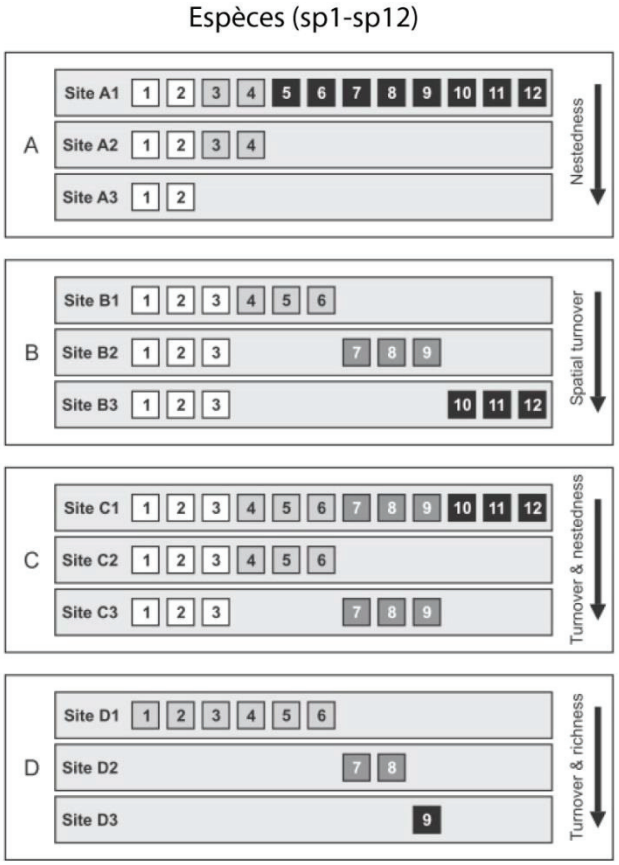


Figure 4 : Décomposition de la béta-diversité en ses composantes de renouvellement spatial (« Turnover » ou « spatial turnover ») et de diversité emboitée (« nestedness »). L'exemple illustre 4 cas (A-D) avec 3 sites par cas. Les carrés et leurs numéros représentent la présence des espèces sur chacun des sites. Dans le cas A, la diversité est complètement emboitée (les sites les plus pauvres sont constitués d'un sous échantillon des espèces des sites les plus riches). Dans le cas B, la diversité béta est représentée uniquement par du renouvellement spatial. Tous les sites ont la même richesse (6 espèces) et trois espèces en commun, mais chaque site comporte 3 autres espèces qui lui sont propres. Le cas C présente une béta-diversité influencée à la fois par du renouvellement spatial et de la diversité emboitée. Enfin, le cas D ne comporte pas de diversité emboîtée mais les sites diffèrent de par leur richesse en espèces : la béta-diversité est uniquement sous l'influence du renouvellement spatial. Ce dernier cas illustre la diversité des situations possibles : leur interprétation a été l'objet de nombreux débats relatifs au choix de différents indices de béta-diversité (Carvalho *et al.* 2011; Podani & Schmera, 2011). D'après Baselga (2010b).

La taille moyenne (ou médiane) des aires de répartition des espèces permet d'étudier plus particulièrement le processus de dispersion (Colwell *et al.* 2004, 2009 ; Arita *et al.* 2008 ; Borregaard & Rahbek, 2010 ; Keith & Connolly, 2013). La taille des aires de répartition est quantifiée grâce à des mesures d'étendue d'aires («Extent of occupancy», EOO) ou des mesures de la surface occupée («Area of Occupancy», AOO). Le degré de corrélation entre ces deux mesures dépend de la porosité des aires de répartition, c'est-à-dire de la manière dont une aire est occupée par une espèce (Gaston & Fuller, 2009).

Dans ce travail, richesse, béta-diversité et aire de répartition sont mesurées en utilisant comme unité de base l'espèce. Toutefois, j'utilise deux méthodes d'identification de l'espèce fondées sur des critères morphologiques et génétiques, méthodes qui essaient d'approcher chacune à leur manière le même concept évolutif de l'espèce (Simpson, 1951 ; Hey, 2006). Par ailleurs, la taille de l'aire de répartition est une métrique qui n'est pas indépendante entre espèces, si l'on admet que deux espèces proches auront des niches similaires. J'ai donc utilisé pour cette métrique des méthodes d'analyse qui permettent de prendre en compte la non indépendance phylogénétique des espèces, les contrastes phylogénétiques indépendants (PIC, Felsenstein, 1985) et la régression phylogénétique des moindres carrés généralisés (PGLS, Martin & Hansen, 1997). Les développements actuels des outils moléculaires rendent possibles l'utilisation des métriques de richesse (Diversité Phylogénétique, PD) et de béta-diversité (Phylobétadiversité) qui permettent de prendre en compte la quantité d'évolution séparant les entités biologiques (Faith, 1992 ; Webb *et al.* 2002, 2008 ; Bryant *et al.* 2008 ; Graham & Fine 2008 ; Cadotte *et al.* 2010). De façon générale, ces métriques permettent une inférence beaucoup plus précise des processus, mais leur utilisation reste pour le moment limitée en raison du manque de données moléculaires.

Liens entre métriques

La richesse est la résultante du chevauchement des aires de répartition et peut de ce fait être corrélée avec la taille moyenne des aires de répartition. Toutefois, cette corrélation dépend de la distribution statistique des aires de répartition des espèces et de leur agencement dans l'espace (Borregaard & Rahbek, 2010). Stevens (1989) propose que le gradient latitudinal de richesse soit la conséquence d'une augmentation de la taille des aires de répartition vers les pôles (règle de Rapoport). Cette proposition implique également que la proportion de la béta diversité due à la composante de diversité emboîtée augmente vers les pôles (Baselga 2010b ; Leprieur et al. 2011). Dans les faits, la relation négative forte entre richesse et taille moyenne des aires de répartition des espèces proposée par Stevens (1989) est loin d'être validée. Weiser *et al.* (2007) montrent que la richesse et la taille des aires de répartition des espèces de plantes ligneuses sont négativement corrélées en Amérique du nord, que cette corrélation s'affaiblit en Amérique du sud et qu'elle n'est plus significative à l'échelle du continent américain . Ici encore, plutôt que souligner l'existence d'un seul et même mécanisme, ces variations de corrélation indiquent très probablement qu'il peut exister plusieurs mécanismes à l'origine d'un même patron.

Les patrons peuvent être biaisés

Un point crucial lorsqu'il s'agit d'expliquer un patron est d'estimer leur robustesse vis-à-vis de deux types de biais : le biais taxonomique (« Linnean shortfall ») et le biais d'échantillonnage («Wallacean Shortfall ») (Diniz-Filho *et al.* 2013). Le biais taxonomique représente une détermination erronée des entités biologiques. Ce biais est notamment révélé par une incongruence entre les entités identifiées par des méthodes morphologiques et moléculaires. Le cas le plus fréquent correspond à la découverte, au sein d'une espèce morphologique, d'un complexe d'entités évolutives génétiquement distinctes appelées espèces cryptiques : on parle alors de diversité cachée (Bickford *et al.* 2007 ; Lefébure *et al.* 2006a ; Trontelj *et al.* 2009). Cette diversité cachée conduit à sous-estimer la richesse, à surestimer potentiellement la taille des aires de répartition, et à sous-estimer la béta diversité. Notons que la diversité cachée n'est pas spécifique aux organismes de petites tailles (Beheregaray & Caccone, 2007 ; Trontelj & Fiser, 2009). Brown *et al.* (2007) ont mis en évidence la présence de 6 entités

évolutives de girafe pouvant vivre en sympatrie, là où morphologiquement une seule espèce comportant plusieurs sous-espèces était reconnue.

Le biais d'échantillonnage est la deuxième source d'incertitude, la variabilité dans l'espace de l'effort d'échantillonnage devenant lui-même une source de variation géographique des métriques (Soberon *et al.* 2007 ; Yang *et al.* 2013). Dans la littérature, les conséquences de ce biais ont surtout été évaluées sur les patrons de richesse spécifique et de nombreuses méthodes ont été proposées afin de l'estimer et/ou de le corriger (Smith & Van Belle 1984 ; Colwell & Coddington, 1994 ; Gotelli & Colwell, 2001 ; Magurran, 2004 ; Chao & Jost, 2012 ; Chao *et al.* 2014). En revanche, les conséquences de ce biais sur les patrons de taille des aires de répartition et de béta diversité restent largement inexplorées.

1.1.3) Les quatre grands processus

Les patrons de biodiversité sont façonnés par quatre forces évolutives les « big four » dont les actions se propagent sur les processus agissant aux niveaux d'organisation biologique supérieurs (cf. Fig. 1 et paragraphe 1.1.1). Toutefois, bien que le nombre de processus soit limité, estimer leur contribution relative dans le façonnement de la biodiversité représente un véritable challenge : c'est le cœur de la boîte noire (Fig. 1). Les inférences sur leur contribution restent le plus souvent très indirectes car elles émanent d'approches corrélatives entre facteurs environnementaux et métriques de la biodiversité. C'est aux niveaux d'organisation biologique les plus bas (gènes, individus, populations) que le cadre théorique définissant les processus a été le mieux appréhendé. Ces processus correspondent aux quatre forces évolutives décrites dans le cadre de l'évolution moléculaire et de la génétique des populations (Kimura, 1983).

- La mutation correspond à toutes les modifications du matériel génétique d'une cellule ou d'un virus par des substitutions, des insertions, des délétions de nucléotides ou des réarrangements de segments chromosomiques. La mutation est la cause première de la variabilité génétique.
- La sélection naturelle correspond à un tri des allèles ou des individus en fonction de leur aptitude à survivre et à se reproduire. La valeur sélective des individus (« fitness »), c'est-à-dire leur capacité à transmettre leurs gènes à la génération suivante, représente une mesure de la sélection naturelle.
- La dérive représente la fluctuation aléatoire des fréquences alléliques dans une population entre générations.
- La migration représente l'échange d'individus entre populations. Elle est la cause du flux génique lorsque ces individus se reproduisent.

L'action conjointe de ces forces évolutives se répercute directement sur la structuration des niveaux d'organisation biologique supérieurs (i.e. plus intégratifs). L'absence de migration entre deux populations d'une même espèce, combinée aux effets quelques fois antagonistes de la sélection naturelle et de la dérive, peuvent favoriser leur isolement reproducteur et à la formation d'une nouvelle espèce.

Les processus façonnant l'assemblage des communautés ont souvent été perçus comme « un bazar » (Vellend, 2010). Historiquement, la répartition des espèces et l'assemblage des communautés ont longtemps été envisagés sous le seul angle de la sélection : c'est la niche des écologistes (Grinnell, 1917, Hutchinson, 1957). La niche est un concept qui au sens Hutchinsonien (1957) représente un hyper-volume dans un espace multidimensionnel dont les axes représentent chacun une

ressource ou une condition de l'environnement abiotique et biotique. Cet hyper-volume fixe les limites au sein desquelles un organisme peut vivre et se reproduire. Plus récemment, la théorie neutraliste de l'assemblage des communautés (Hubbel, 2001) et le concept de méta-communauté (Leibold *et al.* 2004) reconnaissent que la variation aléatoire de l'abondance des individus et la dispersion sont des composantes majeures de la structuration des communautés. Ces théories et concept « neutralistes » s'appuient certes sur une hypothèse non réaliste qui consiste à considérer les espèces comme équivalentes mais elle offre un modèle neutre contre lequel le rôle de la sélection peut être testé (McGill et al. 2006, voir Clark, 2008, pour un débat sur la notion de neutralité). En 2010, Vellend (2010) a tenté d'apporter un peu d'ordre dans l'écologie des communautés en proposant une synthèse conceptuelle analogue à celle développée dans le cadre de la génétique des populations. Cet auteur propose que quatre processus structurent les communautés.

- La spéciation représente la « création » de nouvelles espèces.
- La sélection représente le différentiel de fitness des individus appartenant aux différentes espèces de la communauté. Sachant que la valeur sélective d'un individu représente sa capacité à transmettre ses gènes à la génération suivante, la sélection correspond aux différences du nombre des descendants produit par les individus appartenant aux différentes espèces par unité de temps.
- La dérive écologique représente la fluctuation aléatoire de l'abondance relative des différentes espèces au sein de la communauté.
- La dispersion correspond au mouvement d'émigration ou d'immigration des individus et donc des espèces entre les communautés.

Au sein de cette hiérarchie de processus, la spéciation (ou mutation) est génératrice de nouveauté, en revanche, la sélection, la dérive et la dispersion (migration) vont influencer l'abondance relative des entités biologiques dans la communauté (population). Les fluctuations d'abondance des entités biologiques peuvent conduire à l'extinction des entités en un lieu et à un temps donné. L'extinction peut donc être perçue comme la résultante des processus de sélection, de dérive (écologique ou génétique) et d'une absence de dispersion. De ce fait, la présence ou l'absence d'une espèce dans une communauté est sous la dépendance de trois processus : la création de cette entité (spéciation), la dispersion et l'extinction (Brown 1995 ; Wiens, 2011).

1.1.4) Vers une synthèse des facteurs environnementaux

La contribution relative des processus aux patrons de biodiversité est largement sous l'influence d'un certain nombre de facteurs environnementaux. C'est pourquoi, établir et comprendre l'influence de facteurs environnementaux sur les processus représente le deuxième niveau de complexité dans la compréhension des mécanismes façonnant les patrons de biodiversité (voir boite noire Fig. 1). La très grande diversité des facteurs environnementaux a donné naissance à une quantité impressionnante d'hypothèses pour expliquer la distribution spatiale de la biodiversité, Palmer (1994) dénombre plus de 120 hypothèses. Toutefois, nombre d'entre elles sont redondantes et non indépendantes (Rohde, 1992). Dans un désir de synthèse, les facteurs ont été organisés hiérarchiquement pour ne reconnaître au sommet de la hiérarchie qu'un nombre restreint de facteurs synthétiques (cf. Fig. 1) Ces derniers peuvent varier en nombre selon les auteurs (Rohde, 1992 ; Whitakker *et al.* 2001 ; Rahbek & Graves 2001 ; Field *et al.* 2009) et en fonction du groupe taxonomique d'intérêt. Toutefois, trois sont systématiquement retenus dans la littérature - le climat / l'énergie, l'espace / l'hétérogénétité spatiale et l'histoire. Ils sont détaillés ci-dessous, ainsi qu'un autre

facteur d'importance, les interactions biotiques. Enfin, le rôle des interactions entre facteurs est abordé.

1.1.4.1) Climat/énergie

Les relations générales entre la quantité d'énergie reçue par une communauté biologique et sa diversité sont connues depuis Wallace (1878) et ont été vérifiées depuis chez de très nombreux groupes d'organismes (Rosenzweig, 1995 ; Hawkins *et al.* 2003). Les relations énergie-richesse (Wright, 1983) peuvent être linéaires (macro-échelle) ou uni-modales (micro-échelle) en fonction de l'échelle spatiale (Waide *et al.* 1999 ; Evans *et al.* 2005). Le terme énergie a une portée générale. Toutefois, il est important de distinguer l'énergie ambiante de l'énergie productive, en raison des mécanismes différents qu'elles peuvent impliquer (Fig. 5) (Currie *et al.* 2004 ; Evans *et al.* 2005). Notons que le climat actuel est en grande partie dû à la quantité d'énergie solaire perçue par la terre, de ce fait climat et énergie sont intimement liés et seront abordés dans la suite de ce document uniquement par le terme énergie.

L'énergie ambiante (ou énergie solaire) dépend directement de la quantité de radiation solaire perçue, elle est souvent approximée par la température ou par l'évapotranspiration potentielle (Currie *et al.* 2004). L'énergie ambiante impose des contraintes physiologiques aux organismes qui seront capables ou non de vivre et de se reproduire en fonction de leur tolérance (valence) physiologique (Currie *et al.* 2004). De plus, elle a un impact sur les taux métaboliques et les taux de mutations (Lindgren, 1972 ; Allen *et al.* 2002). La température plus élevée en zone intertropicale pourrait ainsi être l'une des causes du gradient latitudinal de richesse en favorisant la spéciation par l'accélération des taux de mutation (Rohde, 1992).

L'énergie productive correspond à la transformation de l'énergie solaire sous forme organique essentiellement par voie photosynthétique par les végétaux ou les cyanobactéries (Evans *et al.* 2005). Elle est donc intimement liée à la production primaire et dépend de la balance entre énergie solaire et disponibilité en eau (Stephenson 1998 ; Waide *et al.* 1999 ; O'Brien, 2006). Cette énergie organique est accessible sous la forme de ressources trophiques pour les organismes hétérotrophes. Les endroits maximisant la quantité de ressources trophiques disponibles peuvent accueillir plus d'individus permettant aux espèces de maintenir des tailles de populations viables, car moins sensibles à l'extinction (Evans *et al.* 2005). Les zones productives favorisent également la coexistence des espèces ayant des niches trophiques très étroites (« niche breath hypothesis ») et celles spécialisées sur des ressources habituellement rares (« niche position hypothesis »). Une forte énergie productive permet également davantage de niveaux trophiques (« more trophic level hypothesis ») (Bonn *et al.* 2004 ; Evans *et al.* 2005). *In fine*, la limitation de l'extinction et la facilitation de la coexistence augmente la richesse. Toutefois, la forte productivité permet le développement de spécialisations trophiques qui, lorsqu'elles deviennent adaptatives, tendent à diminuer la dispersion et se traduisent par une diminution de la taille des aires de répartition des espèces (Salisbury *et al.* 2012).

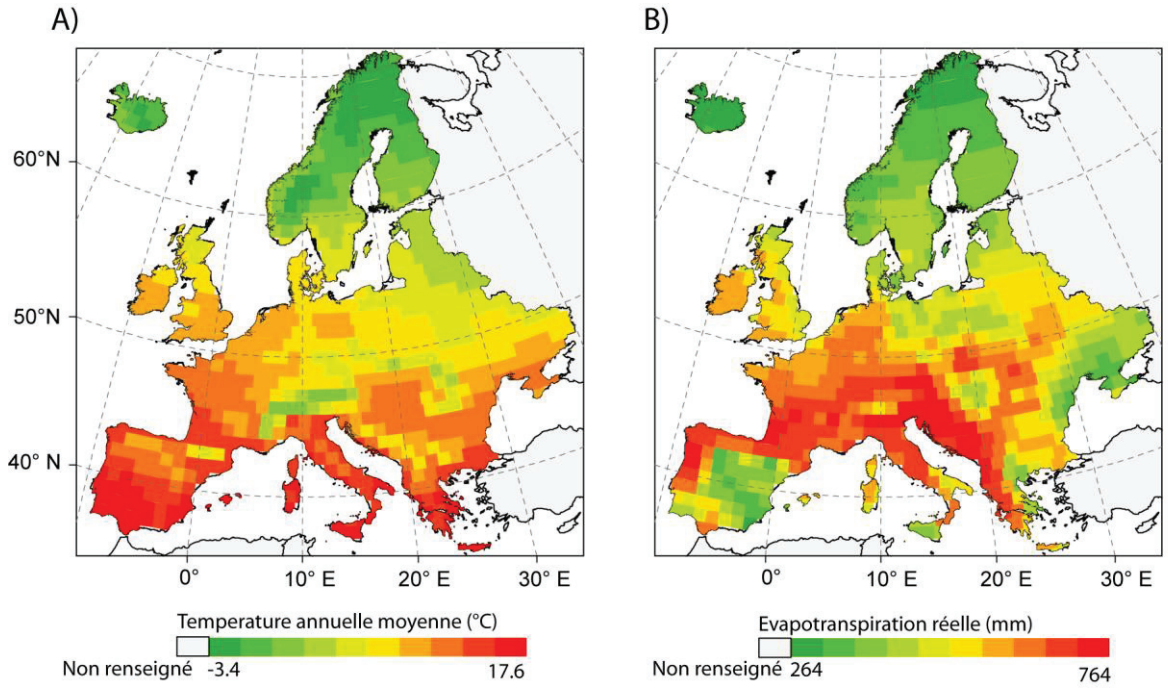


Figure 5 Carte de l'Europe présentant la distribution spatiale des deux composantes de l'énergie : **A)** énergie ambiante représentée par la température annuelle, **B)** l'énergie productive représentée par l'évapotranspiration actuelle.

La variabilité saisonnière de l'énergie, ou saisonnalité, a reçu une attention toute particulière dans la littérature, surtout depuis que Stevens (1989) a proposé qu'elle pouvait être l'une des causes de la règle de Rapoport. Stevens (1989) suggère que les espèces des latitudes élevées sont contraintes de conserver une grande largeur de niche car elles sont confrontées à des variations annuelles d'énergie importantes, notamment de température. Cette large valence écologique favorise la dispersion et l'utilisation de multiples habitats qui participent à l'accroissement de l'aire de répartition des espèces. Au contraire, lorsque la saisonnalité thermique est très peu marquée, comme en zone intertropicale, la sélection favorise l'accroissement des performances sur une gamme thermique étroite (Janzen, 1967). Cette spécialisation représente un frein à la dispersion en limitant les organismes à des habitats dont la température est compatible avec leur valence physiologique étroite (Jocque *et al.* 2010 ; Sunday *et al.* 2010). Le gradient thermique altitudinal au niveau des tropiques représente une barrière d'autant plus infranchissable que les organismes ne sont jamais confrontés à de forts écarts thermiques au cours de la saison au niveau des plaines et vallées (Janzen, 1967 ; Ghalambor *et al.* 2006 ; McCain, 2009). Dès les années 70, cette hypothèse est formulée par Janzen (1967) qui pose la question « Are mountain passes higher in the tropics ? » Cette hypothèse est corroborée par la petitesse des tailles des aires de répartition de la plupart des groupes taxonomiques occupant les chaines de montagnes des zones tropicales, telle que la cordillère des Andes (Buckley & Jetz, 2007 ; Davies *et al.* 2007a ; Ruggiero & Hawkins, 2008).

1.1.4.2) Espace et hétérogénéité spatiale

La surface d'un habitat et son hétérogénéité environnementale (relief, climat) sont deux facteurs connus pour entretenir des liens étroits avec la diversité biologique. Trois explications non mutuellement exclusives supportent les relations surface-richesse (MacArthur & Wilson, 1967) : (1) le taux d'extinction et (2) le taux de spéciation dépendent de la surface (MacArthur & Wilson, 1967 ;

Losos & Schluter, 2000) ; 3) l'hétérogénéité environnementale est corrélée positivement à la surface et favorise la coexistence d'espèces ayant des exigences écologiques variées (Kerr *et al.* 2001). Selon la première explication, le risque d'extinction augmente pour de petites surfaces en raison d'une diminution de la taille des populations. Etant donné que la quantité totale d'énergie productive disponible dépend aussi de la surface, les deux explications sont souvent intrinsèquement liées (Wright, 1983). La deuxième explication suggère que le taux de spéciation augmente avec la surface en raison du processus d'isolement par la distance. Par ailleurs, les espèces occupant des habitats étendus sont confrontées à d'avantage d'hétérogénéité environnementale et de barrières géographiques.

Les hétérogénéités, topographique et d'habitat, deux facteurs souvent confondus dans la littérature (Kerr *et al.* 2001), favorisent les spéciations et limitent l'extinction, ce qui se traduit par une augmentation de la richesse régionale, de la béta diversité et par une diminution de la taille des aires de répartition. Ces hétérogénéités accentuent le processus de spéciations allopatrique et parapatrique en imposant des barrières physiques ou écologiques aux flux de gènes entre populations (Qian & Ricklefs, 2000 ; Rahbek & Graves, 2001 ; Davies *et al.* 2007a ; Dias *et al.* 2013). Les gradients altitudinaux marqués associés aux zones fortement hétérogènes sont reconnus pour jouer un rôle tampon contre l'extinction en servant de zone de survie (refuge) lors de changement climatique (Ohlemüller *et al.* 2008 ; Rull, 2009 ; Stewart *et al.* 2010 ; Keppel *et al.* 2012). Lors des phases d'instabilité climatique, il suffit en effet aux espèces de disperser sur de courtes distances pour maintenir leur valeur sélective.

Enfin, la configuration spatiale de la zone (domaine), en l'absence de tout autre facteur, peut avoir une influence sur la distribution de la diversité par les contraintes géométriques qu'elle impose sur le placement des aires de répartition des espèces. Un modèle connu sous le nom de *Mid Domain Effect* (MDE ; Colwell & Lee, 2000 ; Colwell *et al.* 2004) prévoit une augmentation de la richesse dans la partie médiane d'un domaine (e.g. le centre d'un continent) en raison du chevauchement des aires de répartition des espèces largement distribuées. Ce patron général est plus sensible lorsque la proportion d'espèces largement distribuées est importante (Dunn *et al.* 2007). Lorsqu'une telle proportion diminue, les patrons liés aux effets de domaine deviennent plus complexes (Colwell *et al.* 2009). Le MDE, en plus d'avoir fait l'objet d'un intense débat, a reçu un support mitigé et n'apporte pour le moment que peu d'explication aux gradients de richesse (Zapata *et al.* 2005).

1.1.4.3) L'histoire

Le concept d'histoire en écologie exprime l'idée qu'un patron de biodiversité observé au temps t dépend aussi des évènements intervenus au temps t-1. Il s'ensuit que les patrons ne traduisent pas nécessairement un état d'équilibre entre la biodiversité et les facteurs contemporains (Baselga *et al.* 2012). Ils retiennent aussi l'empreinte des changements climatiques et environnementaux apparus au cours de l'histoire (Leprieur et al. 2011 ; Condamine *et al.* 2013 ; Morueta-Holmes et al. 2013). Les fluctuations climatiques brutales ainsi que les évènements stochastiques de type météoritiques, volcaniques ou géologiques sont souvent associés à une augmentation massive du processus d'extinction. Ils constituent les facteurs principaux à l'origine des cinq grandes crises d'extinction au cours de l'histoire (Arens & West, 2008 ; Condamine *et al.* 2013). Sans générer de crise d'extinction massive, les oscillations climatiques du Pléistocène (les 2.5 derniers millions d'années), alternant phases glaciaires et interglaciaires, ont eu des conséquences profondes sur la répartition de la biodiversité actuelle sur l'ensemble de la planète et plus particulièrement en Europe et en Amérique du

nord (Hewitt, 1996, 2000 ; Schmitt, 2007 ; Svenning & Skov, 2004, 2007 ; Svenning *et al.* 2008 ; Araújo *et al.* 2008 ; Hortal *et al.* 2011 ; Leprieur *et al.* 2011). En règle générale, les phases glaciaires plus marquées aux latitudes les plus élevées (nord de l'Europe, Canada, voir Fig. 6) ont entraîné l'extinction de certaines d'espèces et la migration d'autres espèces vers des refuges (Hewitt, 1996). En fonction des capacités de dispersion des organismes, l'empreinte de ces phases glaciaires sur les patrons actuels de richesse est aujourd'hui plus ou moins visible (Hof *et al.* 2008). Ainsi, des organismes dispersant peu sont susceptibles de ne pas avoir eu le temps de (re)coloniser les zones impactées qui présentent de ce fait, encore aujourd'hui, une richesse plus faible (Svenning & Skov, 2007). En revanche, les zones climatiquement plus stables ont pu servir de refuges pour les organismes, leur permettant de survivre lors de l'avancée des glaces et du permafrost (Hewitt, 1996, 1999, 2000). L'isolement des populations dans des refuges a pu favoriser la différenciation génétique marquée, conduisant quelques fois à des processus de spéciation (Hewitt, 1996). Jansson & Dynesius (2002) ont proposé que la succession des cycles glaciaires au cours du Pléistocène (plus d'une vingtaine) aurait sélectionné des organismes généralistes et des organismes disposant de meilleures capacités de dispersion. *In fine*, cette sélection sur le long terme participe à une augmentation des aires de répartition des espèces et constitue un des mécanismes avancés pour expliquer la règle de Rapoport (Dynesius & Jansson, 2000). Toutefois, comme la plupart des milieux montrent une forte corrélation entre la variabilité climatique à long terme et la saisonnalité, il est difficile de distinguer quelle est l'échelle temporelle de variabilité climatique la plus susceptible de générer un tel patron (Morueta-Holmes *et al.* 2013 ; Veter *et al.* 2013).

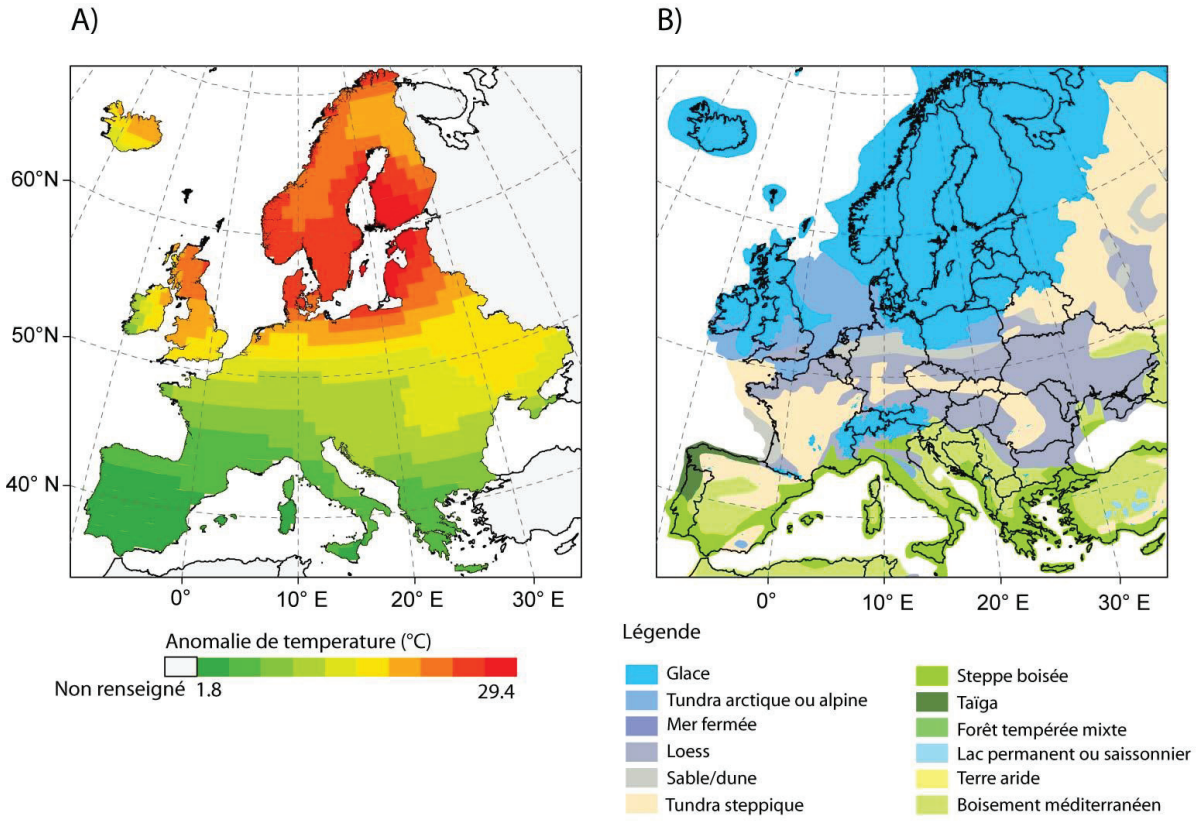


Figure 6 Cartes illustrant les conditions climatiques du dernier maximum glaciaire (-20 000 ans). **A)** L'anomalie de température (différence entre les températures actuelles et celles estimées lors du dernier maximum glaciaire) est un prédicteur de la variabilité climatique au cours du Pléistocène (Araújo *et al.* 2008 ; Leprieur *et al.* 2011). **B)** Carte de la couverture paysagère lors du dernier maximum glaciaire permettant de voir notamment l'étendue des glaciers (en bleu), (carte réalisée à

partir de données de Ehlers *et al.* 2004 ; Hughes & Woodward, 2008, Commision for the Geological Map of the World)

1.1.4.4) Interactions biotiques

Le rôle des interactions biotiques (relation entre les êtres vivants) sur la genèse des patrons de diversité à large échelle est largement inexploré et représente l'un des plus grands challenges à venir (Lavergne *et al.* 2010 ; Thuiller *et al.* 2013 ; Wisz *et al.* 2013). La distribution spatiale des espèces peut être sous l'influence de nombreuses interactions biotiques telles que la prédation, la compétition, les interactions prédateur-proie, les interactions hôte-parasite, la facilitation, et le mutualisme (Wisz *et al.* 2013). Les interactions biotiques intervenant localement, leur impact sur la distribution des espèces a principalement été envisagé à l'échelle locale, et qui plus est, sur un nombre très restreint d'espèces (Pearson & Dawson, 2003 ; Wisz *et al.* 2013 ; voir tout de même Pigot & Tobias, 2013 pour une exception). La complexité des interactions et leur diversité pouvant potentiellement entraîner des conséquences opposées sur la diversité, leurs études sur un cortège d'espèces important à large distribution s'avère rapidement cauchemardesque. Schemske et ses collaborateurs (2009) ont proposé une des rares études de synthèse portant sur l'implication des différentes interactions biotiques sur la répartition de la biodiversité. Leur conclusion suggère que la grande diversité des interactions biotiques dans la zone équatoriale pourrait être l'un des moteurs du gradient latitudinal de richesse, mais que beaucoup reste à faire avant qu'un tel mécanisme puisse être réellement testé.

L'exclusion compétitive des espèces représente certainement l'interaction dont les conséquences ont été les plus étudiées à large échelle spatiale (Webb *et al.* 2002 ; Davies, 2006 ; Pigot & Tobias, 2013). L'exclusion compétitive intervient lorsque des espèces occupent des niches écologiques proches et entrent en compétition pour l'accès aux ressources, sachant que les espèces apparentées ont plus de chance de partager des niches écologiques semblables en raison de l'inertie phylogénétique (Wiens, 2004). Cette compétition se traduit spatialement par une disjonction des aires de répartition des espèces. Par exemple, les aires de répartition très faiblement chevauchantes de deux espèces de hérisson européen (*Erinaceus europeaus, E. roumanicus*) reflètent leur histoire biogéographique et notamment la localisation de leur refuge glaciaire. Cependant, la très faible zone de contact suggère que des mécanismes de compétition expliquent leurs aires disjointes (Santucci *et al.* 1998, voir Fig. 7). De plus, la mise en évidence de zones où les espèces au sein d'une communauté sont plus divergentes génétiquement entre elles (sur-dispersion phylogénétique) qu'attendues sous un modèle nul, suggère que les espèces proches tendent à s'exclure spatialement (Cavender-Bares *et al.* 2004, 2006, 2009 ; Pigot & Tobias, 2013).

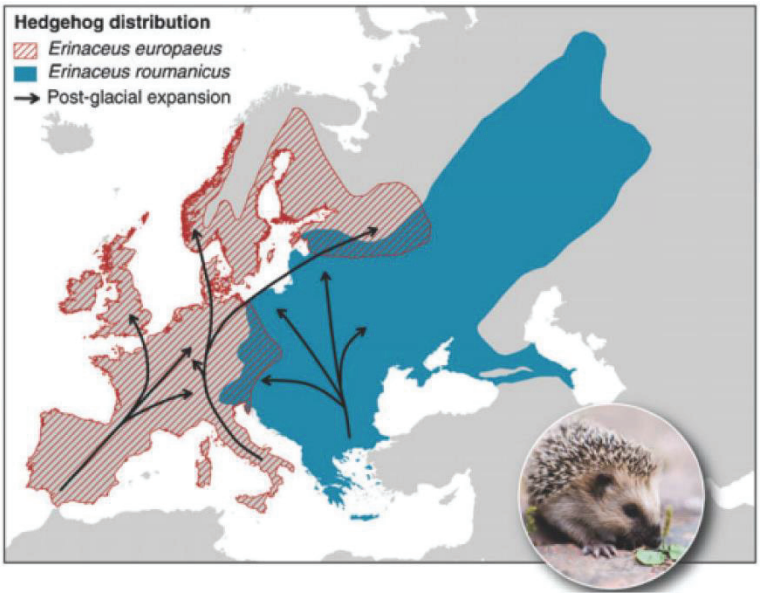


Figure 7 : Aire de répartition des deux espèces de hérisson européen avec les routes d'expansion postglaciaire proposées par Hewitt, (1999). Les aires largement disjointes présentant des zones de contact très réduites proviennent vraisemblablement de compétition ou d'interactions négatives. D'après Wisz *et al.* (2013).

1.1.4.5) Les facteurs peuvent interagir

Les facteurs environnementaux peuvent naturellement interagir, leur résultante ne correspondant plus à l'addition de leurs effets simples. Pourtant, ces interactions entre facteurs sont assez rarement prises en compte par les modèles de diversité en macro-écologie. Certaines études ont montré des interactions positives entre énergie et hétérogénéité spatiale sur la richesse spécifique (Kerr & Packer, 1997 ; Rahbek & Graves, 2001 ; Ruggiero & Kitzberger, 2004). Ainsi les régions montagneuses en zone tropicale disposant d'abondantes ressources trophiques sont caractérisées par une augmentation de richesse à l'échelle régionale (Rahbek & Graves 2001 ; Ruggiero & Hawkins, 2008 ; Davies *et al.* 2007a). En l'absence de quantités suffisantes de ressources trophiques ou d'énergie ambiante, les zones spatialement hétérogènes ne contiennent pas plus d'espèces (Kerr *et al.* 1997).

1.2) Approches pour démêler les liens entre patrons, processus et facteurs

Les difficultés inhérentes à la compréhension des liens entre patrons, processus et facteurs ont favorisé la segmentation des disciplines relatives à l'écologie et à l'évolution. Ces disciplines utilisent des approches et des outils différents et ne disposent pas toutes du même pouvoir d'inférence et de généralisation. Je synthétise brièvement le champ d'action de certaines de ces approches qui sont utilisées dans ce travail.

1.2.1) Raisonner à large échelle spatiale et temporelle : approche macro-écologique

Du mécanisme universel aux principes de mutli causalité et de non stationnarité

Raisonner à large échelle spatiale et temporelle sur des cortèges importants d'espèces a pour objet de fournir une compréhension sur les liens entre processus et facteurs environnementaux qui soit emprunte de généralités (Fig. 8). Une telle démarche est à l'origine de l'approche macro-écologique (Brown & Maurer, 1989). Brown (1995) définit la macro-écologie comme « une investigation statistique non expérimentale des relations entre la dynamique des populations d'espèces qui ont été typiquement étudiées à de petites échelles par les écologues et les processus de spéciation, d'extinction, et d'expansion ou de contraction des aires de répartition qui ont été étudiés à de plus grandes échelles par les biogéographes, les paléontologues et les macro-évolutionnistes ».

L'approche macro-écologique est dans sa définition assez récente (Brown & Maurer, 1989), même si de nombreuses études relevaient déjà de cette discipline avant sa définition. C'est notamment le cas de l'analyse des gradients latitudinaux de richesse proposée par Pianka (1966). A ces débuts, l'approche macro-écologique est souvent une quête pour un mécanisme unique et universel (cf. paragraphe 1.1.1). Par exemple, la théorie métabolique cherche à expliquer le gradient latitudinal de richesse uniquement à partir des effets de la température sur le taux métabolique (Allen *et al.* 2002 ; Brown *et al.* 2004), mais cette théorie n'est pas validée (Hawkins *et al.* 2007). Par la suite, l'émergence d'explications alternatives non mutuellement exclusives aboutit à reconnaître le principe de multi-causalité (Whitakker *et al.* 2001). Cette reconnaissance a des impacts conceptuels importants et se traduit notamment par le développement d'une exigence méthodologique propre à la macro-écologie. Il ne s'agit plus de tester indépendamment chacune des hypothèses vraisemblables mais bien d'évaluer conjointement l'influence relative des différentes hypothèses dans la genèse des patrons de biodiversité (Tisseuil *et al.* 2013 ; Gouveia *et al.* 2013). Plus récemment, le principe de non stationnarité est venu s'ajouter à celui de multi-causalité car il apparaît que l'influence relative des différents facteurs varie dans l'espace (Foody, 2004 ; Svenning *et al.* 2009 ; Hortal *et al.* 2011). Des approches analytiques récentes issues de la géographie statistique permettent actuellement de tester la variation spatiale de l'influence relative des différents mécanismes (Brundson *et al.* 1996 ; Fortheringham *et al.* 2002 ; Eiserhardt *et al.* 2011 ; Gouveia *et al.* 2013).

Les limites des approches corrélatives entre métriques et facteurs

La très grande majorité des études menées sur d'importants cortèges d'espèces infèrent les processus à partir de corrélations entre métriques de biodiversité et facteurs environnementaux. Cette approche pose deux problèmes. Premièrement, la corrélation même forte n'implique pas nécessairement un lien de causalité. Deuxièmement, les processus ne sont pas explicitement testés : ils sont seulement déduits *a posteriori* à partir de l'identification des effets de différents facteurs. Par exemple, une corrélation négative entre la diversité et l'intensité des glaciations du Pléistocène est généralement interprétée comme étant la conséquence d'une extinction forte et d'une absence de recolonisation due à des contraintes à la dispersion (Svenning & Skov, 2007 ; Araújo *et al.* 2008 ; Hof *et al.* 2008 ; Svenning *et al.* 2008). Face aux limitations des approches corrélatives, certains auteurs ont proposé de tester l'influence des processus à partir d'approches mécanistiques qui modélisent explicitement la dispersion (Colwell *et al.* 2004 ; Storch *et al.* 2006 ; Rahbek *et al.* 2007 ; Rangel *et al.* 2007; Davies *et al.* 2007a ; Gotelli *et al.* 2009) ainsi que les processus de spéciation et d'extinction

(modèle Geophylétique ; Brayard *et al.* 2005 ; Escarguel *et al.* 2008). Bien que très prometteurs, ces modèles sont souvent limités par le nombre très restreint de facteurs environnementaux qu'ils peuvent intégrer et par le manque de connaissances théoriques sur les fonctions de lien entre facteurs et processus.

Vers une meilleure inférence des processus pour de multiples espèces

Inférer plus directement le rôle des processus de spéciation, d'extinction et de dispersion est un des objectifs de la macroévolution à travers l'étude de la diversification (différence entre spéciation et extinction) et de la biogéographie phylogénétique à travers la reconstruction des aires de répartition au cours du temps et le test des scénarios de dispersion/vicariance (Morrone & Crisci, 1995 ; Weir & Schluter 2007 ; Morlon *et al.* 2010 ; Ronquist & Sanmartin, 2011 ; Condamine *et al.* 2012). Paradoxalement, les études de diversification et de biogéographie phylogénétique infèrent directement l'influence des processus à partir de registres fossiles ou de phylogénies mais le rôle des facteurs environnementaux est déduit à postériori sans nécessairement être testé statistiquement. L'étude du registre fossile qui a toutefois permis de mettre en évidence plusieurs crises d'extinction (cinq crise majeures au total), de grandes radiations évolutives (angiospermes, mammifères…) et des événements majeurs de dispersion au cours des temps géologiques (Condamine *et al.* 2013 ; Reis et al. 2014). Plus récemment, les développements d'outils moléculaires et analytiques ont permis l'étude de la diversification à partir de phylogénies moléculaires de taxons actuels (Morlon *et al.* 2010 ; Pyron & Burbrink, 2013 ; Morlon 2014). Ces phylogénies lorsqu'elles sont datées grâce à des horloges moléculaires permettent de replacer dans le temps les évènements majeurs de diversification, d'extinction ou de dispersion (Ronquist & Sanmartin, 2011 ; Condamine *et al.* 2012). Les radiations évolutives chez les Papilionidae apparaissent principalement au début de l'Eocène (-55 millions d'années) en raison des climats chauds qui leur auraient permis de coloniser l'Eurasie par le détroit de Béring et de se diversifier sur ce nouveau territoire (Condamine *et al.* 2012).

Les études de diversification à partir de phylogénies moléculaires se limitent encore à un nombre de taxons restreint et sont à ce jour, en l'absence de registre fossile, encore limitées par la modélisation des taux d'extinction (Morlon, 2014). Les approches de diversification et de biogéographie phylogénétique sont particulièrement prometteuses mais elles souffrent pour le moment de plusieurs limitations :

- Hormis une étude très récente incorporant l'effet de la température sur les taux de diversification (Condamine *et al.* 2013), l'effet des facteurs environnementaux n'est pas explicitement pris en compte dans les analyses, ce qui ne permet pas de quantifier leur influence relative sur les processus (cependant voir Roquet *et al.* 2013a et Struwe *et al.* 2011).
- Pour le moment, ces études examinent des temps longs et négligent les processus populationnels.
- Les phylogénies datées souffrent d'une forte imprécision de datation en l'absence d'un registre fossile suffisamment abondant (Penny, 2005 ; Ho *et al.* 2008).
- Les études de biogéographie phylogénétique utilisant des phylogénies pour reconstruire les aires de répartition au cours du temps et tester des scénarios de dispersion/vicariance conservent une résolution spatiale assez grossière et un nombre limité d'entités spatiales (mais voir Landis *et al.* 2013 pour un contre-exemple récent).

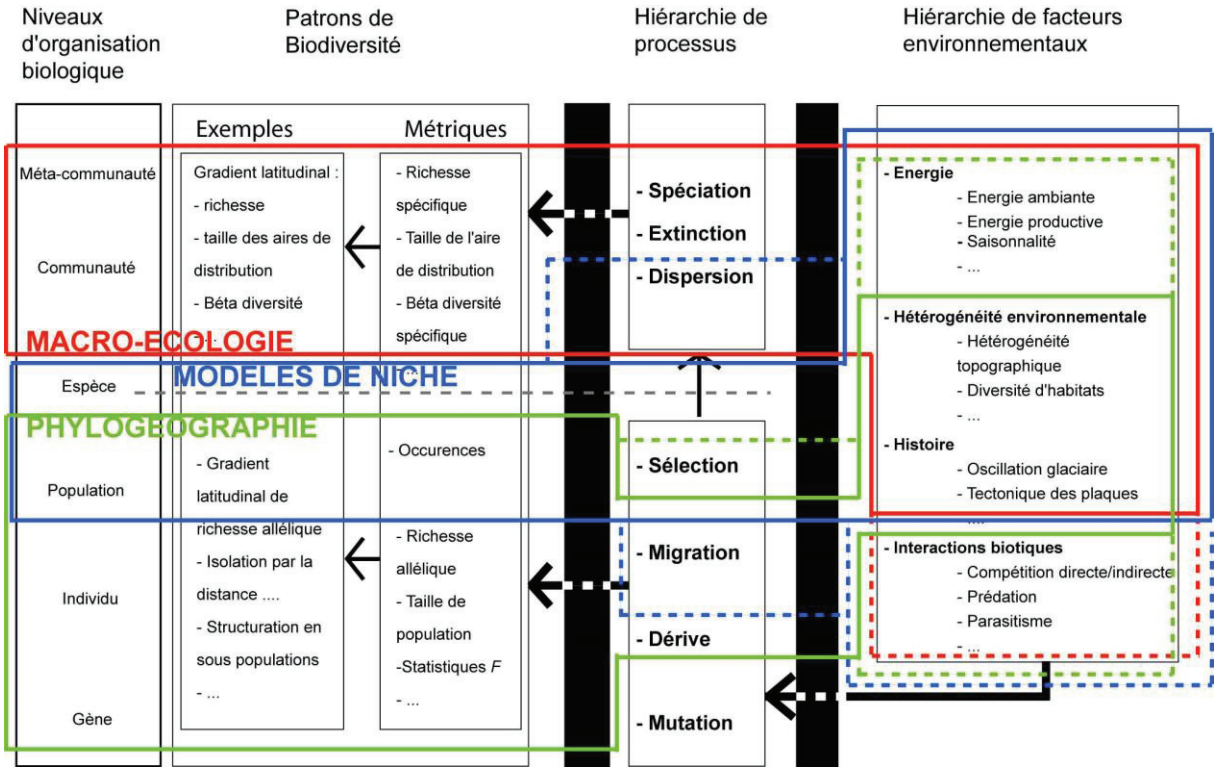


Figure 8 : Schéma conceptuel présentant le positionnement des champs de la macro-écologie (rouge), des modèles de niche (bleu) et de la phylogéographie (vert) dans la compréhension des liens entre patrons de biodiversité, processus et facteurs environnementaux (cf. Figure 1). Les traits continus de couleur montrent les inférences classiquement réalisées par les différents champs. Les traits pointillés de couleur suggèrent les possibilités d'inférences rarement explorées mais en pleine émergence.

1.2.2) Perdre en généralité mais se rapprocher des liens entre processus et facteurs

Se rapprocher des forces évolutives (mutation, sélection, dispersion, dérive) et des liens qu'elles entretiennent avec les facteurs environnementaux pour dicter la répartition spatio-temporelle de la biodiversité, requiert de travailler à des niveaux d'organisation biologique inférieurs à celui de l'espèce. Pour ce faire, d'autres approches sont utilisées mais ce changement se réalise au détriment de la généralité apportée par les approches macro-écologiques.

1.2.2.1) Emphase sur la sélection par les approches de niches

La sélection exercée par les facteurs environnementaux est certainement le processus le plus étudié quand il s'agit de comprendre la distribution de la biodiversité. En l'absence de toutes contraintes, l'espèce occupe une aire de répartition correspondant à sa niche fondamentale. Sous l'effet de contraintes à la dispersion ou d'interactions biotiques négatives, l'espèce occupe une aire plus restreinte, plus étroite, correspondant à sa niche réalisée. C'est cette dernière qui est déterminée lorsque le chercheur s'appuie sur des données distributionnelles (Pulliam, 2000). Toutefois, l'aire occupée peut s'étendre au-delà des limites imposées par la niche fondamentale en raison d'une

dynamique source-puits ou du temps nécessaire au processus d'extinction (Pulliam, 1988 ; Oberdorff *et al.* 2011).

Les modèles de niche ou de distribution spatiale (SDM, species distribution modelling) offrent la possibilité d'étudier l'impact du processus de sélection exercé par les facteurs environnementaux sur la distribution d'une espèce (Fig. 8). Ils permettent d'établir des liens statistiques entre des données d'occurrence des espèces (présence, ou présence /absence ou abondance) et des facteurs environnementaux (Guisan & Zimmermann, 2000). La plupart des modèles étudient l'influence de trois types de facteurs (Guisan & Thuiller, 2005) : i) les facteurs de régulation (e.g. température, pH) qui limitent la présence des espèces en raison de leur tolérance physiologique, ii) les facteurs de perturbation qui affectent l'habitat des espèces (naturels ou anthropiques), iii) les facteurs de ressources qui correspondent à ce qui est assimilé par l'organisme (énergie productive, eau…). La modélisation se fonde sur l'hypothèse que les espèces sont en équilibre avec leur environnement (Nogues-Bravo, 2009), ce qui est probablement erroné pour un grand nombre d'entre elles en raison des contraintes à la dispersion (Costa *et al.* 2008 ; Baselga *et al.* 2012). Dans la mesure où ces modèles, pour la plupart corrélatifs, estiment la niche réalisée de l'espèce (Guisan *et al.* 2002 ; Kearney, 2006), ils peuvent s'avérer relativement imprécis lorsqu'ils sont utilisés pour prédire la distribution des espèces en réponse à des modifications de l'environnement (Guisan *et al.* 2002 ; Thuiller *et al.* 2013). De plus, ils ne prennent pas en compte l'évolvabilité des individus vis-à-vis de l'environnement, c'est-à-dire leurs capacités à développer des adaptations et donc de modifier leur niche fondamentale (Lavergne *et al.* 2010 ; Boucher *et al.* 2012 ; Roquet *et al.* 2013a). Très récemment des modèles de niche corrélatifs ont tenté d'incorporer le processus de dispersion, la possibilité d'évolution de la niche, et l'action des interactions biotiques entre espèces (Kearney & Porter, 2009 ; voir Thuiller et al. 2013).

Une des solutions pour estimer le décalage entre niche réalisée et niche fondamentale est de tester expérimentalement au laboratoire la gamme de tolérance des organismes vis-à-vis des principaux facteurs de régulations ou trophiques. Ces tests de laboratoire permettent également de révéler des phénomènes d'adaptation locale chez des populations développant une spécialisation vis-à-vis de leur environnement local (Kawecki & Ebert, 2004 ; Blanquart *et al.* 2013). Ces tests expérimentaux constituent les entrées de certains modèles de niche mécanistiques (biophysiques) qui restent malgré tout peu employés en raison de l'importance de la quantité d'informations physiologiques et comportementales requise (Kearney & Porter, 2009). Ces approches mécanistiques sont donc réservées aux espèces très bien étudiées (Morin & Thuiller, 2009 ; Lauzérale, 2012).

1.2.2.2) Vers une prise en compte de la dispersion et de la dérive par la phylogéographie.

Les mouvements des organismes dans l'espace et la fluctuation démographique aléatoire des individus (et donc des allèles) constituent deux processus clés dans la genèse des patrons de diversité génétique des populations. De tels patrons sont étudiés par la phylogéographie. Avise (2000) définit ce champ disciplinaire comme l'étude des principes et des processus gouvernant la distribution géographique de lignées généalogiques au sein et entre espèces proches. Autrement dit, la phylogéographie intègre les dimensions spatiales et temporelles de la généalogie (Avise, 2009) et base son étude sur l'inférence de la dynamique spatiale et/ou démographique des populations d'une espèce ou de quelques espèces proches à partir de marqueurs moléculaires (Fig. 8). La phylogéographie a longtemps été très descriptive. Les résultats des reconstructions phylogénétiques et des études de

génétique des populations étaient interprétés dans un cadre géographique (avec une carte) sans que les relations entre patrons et facteurs ne soient explicitement incorporées dans l'analyse.

Depuis le début des années 2000, la phylogéographie a connu deux développements majeurs qui ont permis de prendre en compte dans un cadre analytique statistique explicite et rigoureux le rôle de la dérive et de la dispersion. La prise en compte de la dérive a été possible par l'incorporation de la théorie mathématique de la coalescence (Kingman 1982a,b ; Kuhner, 2008). Elle décrit le processus fortement aléatoire de fusion binaire de tous les lignages d'un échantillon de gènes jusqu'à leur plus proche ancêtre commun (Wakeley, 2009). La théorie de la coalescence établit des relations parfois très complexes entre arbres de gènes, démographies de population au cours du temps, et temps de divergence entre individus (Nielsen & Beaumont 2009). Il est désormais possible d'estimer l'évolution de la taille efficace des populations au cours du temps, de dater des goulots d'étranglement (perte de diversité) ou des croissances démographiques souvent synonyme d'expansion géographique (Excoffier, 2004 ; Drummond *et al.* 2005 ; Heled & Drummond, 2008 ; Excoffier *et al.* 2009 ; Minin *et al.* 2008 ; Gill *et al.* 2013). La deuxième avancée majeure est l'incorporation analytique du contexte géographique de l'étude (Kidd & Ritchie 2006; Lemmon & Lemmon, 2008; Chan *et al.et al.* 2011). Cette incorporation aboutit à estimer directement les flux de gènes entre entité spatiale (nombre de migrants), l'origine géographique d'un clade, les vitesses de dispersion et la dynamique de colonisation spatiale des aires de répartition au cours du temps (Hey & Nielsen, 2007; Lemmon & Lemmon, 2008 ; Hey 2009 ; Lemey *et al.et al.* 2009, 2010 ; Chan *et al.* 2011).

Deux grands types d'approches pour l'étude de la dispersion, tout en incorporant la dérive via la coalescence, ont été développés : les modèles de simulation de données génétiques (Fagundes *et al.* 2007 ; Richards *et al.* 2007 ; Beaumont *et al.* 2009) et les modèles de diffusion spatiale (Lemey *et al.* 2009, 2010 ; Pybus *et al.* 2012). Les modèles de simulation reconstruisent des généalogies attendues sous différents scénarios biogéographiques alternatifs pour les confronter aux données empiriques (Richards *et al.* 2007 ; Nielsen & Beaumont 2009 ; Beaumont *et al.* 2010 ; Csillery *et al.* 2010). Ces méthodes largement implémentées dans un cadre analytique flexible d'Approximate Bayesien Computation (ABC ; Beaumont *et al.* 2002, 2009 ; Csillery *et al.* 2010) sont actuellement en plein développement (Fagundes *et al.* 2007 ; Carsten et Richards 2007 ; Carnaval *et al.* 2009 ; Chan *et al.* 2011 ; Lorenzen *et al.* 2011). Toutefois, ces modèles restent lourds à mettre en place spécialement lorsque le nombre de populations est important car les modèles alternatifs atteignent rapidement une grande complexité (Nielsen & Beaumont, 2009). Les modèles de diffusion spatiale s'appuient sur les approches de reconstruction de caractères ancestraux (Lemey *et al.* 2009, 2010 ; Pybus *et al.* 2012). Ils n'infèrent pas explicitement l'histoire spatiale des populations, mais documentent l'histoire ancestrale de l'échantillon des individus (Bloomquist *et al.* 2010). Ces modèles développés dans un cadre Bayésien reconstruisent simultanément la généalogie, l'histoire démographique et la dispersion sous un modèle de coalescence et de diffusion (Lemey *et al.* 2009, 2010). Ils permettent de reconstruire la dynamique de colonisation de l'aire de répartition du clade au cours du temps afin notamment d'estimer son lieu d'origination, sa vitesse de colonisation, et de dater sa présence en des lieux particuliers (refuges glaciaires…). Ces méthodes sont actuellement employées sur une gamme de plus en plus vaste «d'organismes» allant des virus (Lemey *et al.* 2010 ; Allicock *et al.* 2012 ; Faria *et al.* 2012 ; Pybus *et al.* 2012) aux ours polaires (Edwards *et al.* 2011). Des implémentations en phylogénie linguistique ont même été développées pour retracer l'origine des langues indo-européennes (Bouckaert *et al.* 2012).

Afin de gagner en généralité, ces approches de phylogéographie sont menées dans un cadre comparatif en utilisant plusieurs taxons. Toutefois, les inférences sont la plupart du temps spécifiques à chaque taxon pris individuellement (Hewitt, 1996 ; Taberlet *et al.* 1998; Lorenzen *et al.* 2011). Le

développement récent d'une approche de simulations d'ABC hiérarchiques (HABC ; Hickerson & Meyer, 2008) permet de tester simultanément des scénarios de divergence et de colonisation au cours de la même analyse (Hickerson *et al.* 2006 ; Hickerson & Meyer, 2008 ; Carnaval *et al.* 2009 ; Bell *et al.* 2011 ; Chan *et al.* 2011). Cette approche hiérarchique étant très complexe à mettre en place et dévoreuse de ressources informatiques, elle se limite à un faible nombre de taxons. Le défi d'une phylogéographie comparative au niveau des communautés afin de comprendre la manière dont elles se sont assemblées au cours du temps est encore à relever (Hickerson *et al.* 2010). Dans le chapitre 3 de ce travail, les modèles de diffusion spatiale sont utilisés afin de comprendre les caractéristiques de la dispersion et de reconstruire la dynamique des aires de répartition de plusieurs espèces d'isopodes souterrains.

Vers une prise en compte de la sélection, de la dispersion et de la dérive en couplant approches de niche et phylogéographie.

Les approches de niche et de phylogéographie peuvent être utilisées conjointement pour étudier l'influence relative des processus de sélection, de dérive et de dispersion. Ainsi, l'historique des occurrences d'une espèce peut être inférée par la phylogéographie puis testée à partir d'une reconstruction de la paléo-distribution de l'espèce qui s'appuie sur une connaissance de sa niche actuelle (Cordellier & Pfenninger 2009 ; Barlow *et al.* 2011). Un tel couplage révèle des contraintes à la dispersion liées notamment à la présence de barrières. Un couplage similaire entre modèles de niche et phylogéographie consiste à modéliser la paléo-distribution afin d'élaborer des scénarios alternatifs de la dynamique spatiale de l'aire de répartition d'une espèce au cours du temps. La pertinence de ces scénarios est ensuite évaluée grâce à une simulation de données génétiques et une approche ABC (Richards *et al.* 2007 ; Lorenzen *et al.* 2011 ; Chan *et al.* 2011). Ce second type de couplage offre l'avantage d'évaluer le support relatif des scénarios alternatifs et relaxe les hypothèses réalisées par les modèles de niche (Csillery *et al.* 2010).

Ces approches couplées offrent des perspectives intéressantes mais elles reposent dans leur grande majorité sur la modélisation de la niche réalisée. Autrement dit, l'inférence des processus de sélection et de dispersion ne proviennent pas de mesures indépendantes. Des études proposant une estimation indépendante de ces deux processus restent à ma connaissance tout à fait marginales (Moritz *et al.* 2012 ; Cavender-Bares *et al.* 2011). Il est pourtant possible de façon indépendante d'estimer la niche réalisée, la niche fondamentale et le processus de dispersion en couplant, i) des tests physiologiques en laboratoire, ii) une modélisation de la niche réalisée, et iii) une approche phylogéographique. Un tel couplage est proposé au chapitre 4 de ce travail afin de démêler le rôle relatif de la dispersion et de la sélection sur la répartition d'un isopode aquatique souterrain.

1.3) Faune du milieu aquatique souterrain, un modèle original et nouveau en macro-écologie

1.3.1) Spécificités du milieu aquatique souterrain

Le milieu aquatique continental souterrain offre une stabilité saisonnière des paramètres environnementaux et une simplification des communautés et des réseaux trophiques. Ces caractéristiques le rendent pertinent pour démêler les liens entre patrons de biodiversité, facteurs et

processus. Contrairement à de nombreux habitats de surface, les variabilités thermiques journalière et saisonnière sont extrêmement réduites dans le milieu souterrain. La variabilité thermique décroit de façon exponentielle avec la profondeur, si bien que la température des eaux souterraines conserve au cours des saisons une température proche de la moyenne annuelle de la température de l'air (Freeze & Cherry, 1979). En revanche, à l'image des milieux de surface, le milieu souterrain subit également les grands cycles climatiques tels que les oscillations glaciaires du Pléistocène. De ce fait, il permet d'étudier l'influence des oscillations climatiques à long terme tout en s'affranchissant de la co-variation avec la variabilité thermique saisonnière (Morueta-Holmes *et al.* 2013 ; Veter *et al.* 2013).

La diversité biologique dans les milieux aquatiques souterrains est plus faible que dans les milieux de surface, ce qui limite la complexité des interactions entre organismes. De nombreux groupes taxonomiques sont représentés dans les eaux souterraines (Stoch, 1995 ; Lefébure, 2005 ; Deharveng *et al.* 2009), mais parmi les métazoaires, les crustacés dominent largement les communautés : ils représentent entre 65 à 70% de la diversité (Ferreira *et al.* 2007 ; Deharveng *et al.* 2009; Niemiller & Zigler 2013). Le succès évolutif des crustacés en milieu souterrain est tel que le nombre d'espèces strictement inféodés aux eaux souterraines est identique voire supérieur au nombre d'espèces dans les milieux de surface (Ferreira, 2005, Stoch & Galassi, 2010).

D'un point de vue morphologique, les organismes strictement inféodés aux eaux souterraines, c'est-à-dire y réalisant la totalité de leur cycle de vie, partagent plusieurs traits biologiques qualifiés de troglomorphiques. Les plus communs sont la dépigmentation et la régression des yeux qui conduit à une anophtalmie (Culver *et al.* 1995, voir Fig. 9). Ces organismes présentent également un certain nombre de convergences physiologiques (métabolisme ralenti, résistance au jeûne et à l'hypoxie), de traits d'histoire de vie semblables (une longévité accrue, une fécondité réduite, des œufs plus gros, une maturité sexuelle plus tardive) et un régime trophique supposé omnivore (Ginet & Decou, 1977 ; Malard & Hervant, 1999 ; Gibert & Deharveng, 2002 ; Culver, 2005 ; Lefébure, 2005).

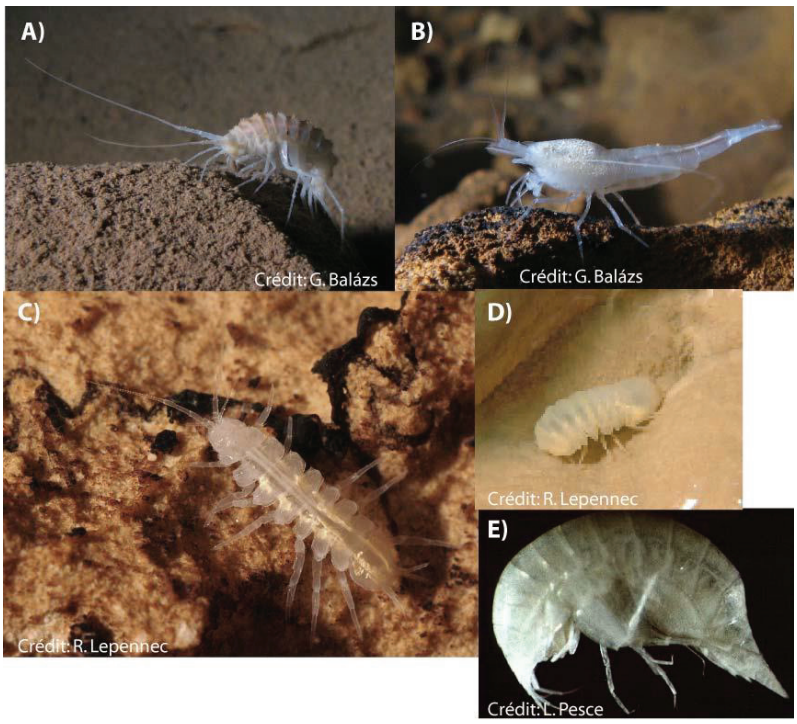


Figure 9 : Exemples de crustacés aquatiques souterrains illustrant les convergences morphologiques (dépigmentation, anophtalmie) entre taxons, **A)** *Niphargus balkanicus* (Absolon, 1927), **B)**

Troglocaris (Speleocaris) pretneri, (Matjašič, 1956), **C)** *Proasellus valdensis*, (Chappuis, 1948), **D)** *Caecosphaeroma virei* Dollfus, 1896, **E)** *Salentinella gracilima* Ruffo, 1947.

La grande majorité des organismes occupe des aires de répartition de petite taille (Gibert & Deharveng, 2002 ; Danielopol *et al.* 2003 ; Christman *et al.* 2005 ; Gibert *et al.* 2009). Nombre d'espèces ne sont connues que d'un seul aquifère dont l'étendue géographique ne dépasse pas quelques kilomètres carrés (Sket, 1981). Théoriquement, la distribution de la biodiversité est donc peu sensible aux contraintes géométriques du *Mid Domain Effect* (Colwell & Lee, 2000 ; Colwell *et al.* 2004 ; Dunn *et al.* 2007). Cet endémisme prononcé pourrait se renforcer au fur et à mesure de la découverte d'une importante diversité cachée mis en évidence par de nombreuses études moléculaires (Lefébure *et al.* 2006a, 2007 ; Finston *et al.* 2007 ; Trontelj *et al.* 2009 ; Zaksek *et al.* 2007, 2009 ; Abrams *et al.* 2012 ; Morvan *et al.* 2013). En effet, la découverte d'espèces cryptiques au sein d'entités morphologiques largement distribuées peut entraîner une diminution conséquente de l'aire de répartition moyenne des espèces. Trontelj et ses collaborateurs (2009) ont ainsi suggéré que les espèces morphologiques dont les aires de répartition s'étendaient au-delà de 200 km comprenaient certainement un complexe d'espèces cryptiques présentant des distributions plus étroites. Les résultats obtenus dans le cadre de ce travail (article 4) suggèrent que les effets de la diversité cachée sur la taille des aires de répartition des espèces souterraines sont en fait bien plus complexes puisqu'elle varie régionalement.

Enfin, le milieu souterrain dispose de quantité de ressources trophiques limitées à l'origine d'une simplification des réseaux trophiques (Gibert & Deharveng, 2002). En effet, l'absence de lumière empêche toute production primaire d'origine photosynthétique et la production primaire liée à de la chimio-autotrophie (production de composés organiques en l'absence de lumière par oxydation de substances inorganiques) est généralement très insuffisante pour maintenir seule les réseaux trophiques (Engel, 2005). De ce fait, les réseaux trophiques sont largement hétérotrophes et sont directement sous la dépendance de l'apport de matière organique en provenance de la surface. Ces apports s'effectuent majoritairement sous une forme dissoute, les flux de carbone organique dissous étant utilisés par les biofilm microbien qui constitue la principale source d'énergie pour les organismes des niveaux trophiques supérieurs (Simon *et al.* 2003 ; Foulquier *et al.* 2011).

1.3.2) Patron de biodiversité en milieu souterrain

La description des patrons de biodiversité en milieu aquatique souterrain est un travail largement inachevé, (Stoch, 1995 ; Ferreira *et al.* 2007 ; Deharveng *et al.* 2009 ; Niemiller & Zigler 2013). Le présent travail contribue à l'inventaire de la biodiversité souterraine à travers la construction de la première base de données d'occurrences à l'échelle européenne (cf. article 2).

La diversité biologique des eaux souterraines a essentiellement été documentée en Europe, en Amérique du nord et en Australie. Très peu de données sont actuellement disponibles pour les tropiques, que ce soit en Amérique du Sud ou plus encore en Afrique et Asie du sud-est (Botosaneanu, 1986 ; Marmonier *et al.* 1993 Jubertie & Decu, 1994 ; Deharveng, 2005 ; Trajano & Bichuette, 2010 ; Brancelj *et al.* 2013). Les premiers patrons de richesse spécifique documentés à l'échelle du continent Européen par Hof *et al.* (2008) et Stoch & Galassi (2010) suggèrent que la richesse spécifique décroît de manière monotone avec la latitude (Fig. 10).Toutefois, ces études utilisent des résolutions spatiales extrêmement grossières (écorégions ou limite administrative des pays). Bien que menée en 2008,

l'étude de Hof *et al.* (2008) s'appuie sur des données issues des travaux de Illies datant de 1978. Or, depuis cette date un très grand nombre d'espèces ont été décrites. De ce fait, le patron décrit par Hof *et al.* (2008) est incomplet voire indicatif et requiert une mise à jour importante (Ferreira, 2005).

En Europe, les données disponibles sur la distribution des espèces à une résolution spatiale plus fine proviennent essentiellement du programme européen PASCALIS (Protocol for the Assessment and Conservation of Aquatic Life In the Subsurface, Gibert *et al.* 2005). L'objectif premier de ce programme consistait à décrire les patrons de biodiversité en utilisant un protocole d'échantillonnage standardisé parmi cinq pays (Belgique, Espagne, France, Italie, et Slovénie). Les études utilisant des résolutions spatiales fines (régions ou cellule de 20*20 km) suggèrent un patron latitudinal plus complexe, présentant une crête de diversité aux latitudes comprises entre 42 et 46° Nord de latitude. Cette crête suggérée aussi bien pour les organismes souterrains terrestres (Culver *et al.* 2006) qu'aquatiques (Michel *et al.* 2009), longerait les massifs montagneux des Cantabriques, des Pyrénées, du Massif central, et des Alpes (Figure 10).

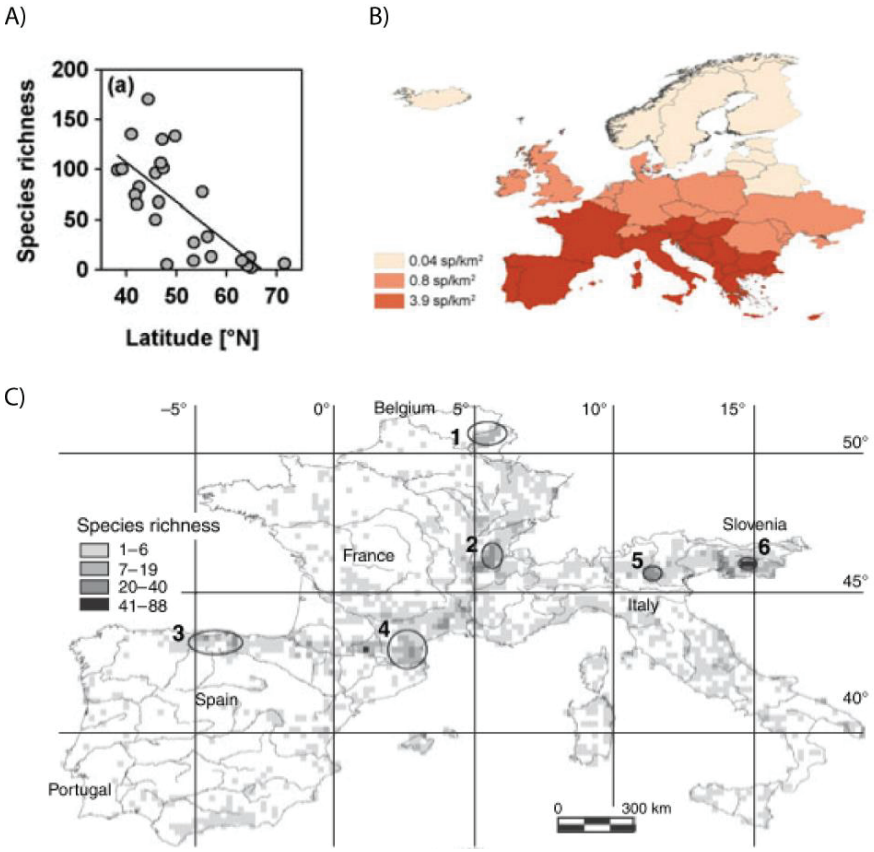


Figure 10 : Patrons géographiques de la richesse spécifique dans les eaux souterraines en Europe. **A)** Patron latitudinal établi à partir de la richesse des 25 écorégions définies par Illies (1978), d'après Hof *et al.* 2008, **B)** Regroupement des pays européens en fonction de leur richesse en crustacés souterrains, d'après Stoch & Galassi, (2010), **C)** Carte de richesse spécifique dans les eaux souterraines de cinq pays européens (Programme, Européen PASCALIS, Gibert *et al.* 2005) utilisant une maille carrée 0.2×0.2°, d'après Michel *et al.* (2009).

1.3.3) Débat sur le rôle de la dispersion et de l'histoire

Le débat entre vicariance et dispersion

Les études de biogéographie relatives au milieu souterrain ont placé le rôle du processus de dispersion au centre du débat dans la genèse des patrons de biodiversité. Partant du postulat que la taille de l'aire de répartition d'un organisme reflète ses capacités de dispersion, les toutes petites aires de répartition des organismes cavernicoles ont longtemps été interprétées comme la résultante d'une dispersion très limitée liée à la fragmentation du milieu (Valentine, 1932 ; Sket, 1981 ; Gibert & Deharveng, 2002 ; Christman *et al.* 2005 ; Culver *et al.* 2009 ; Gibert *et al.* 2009). Cette idée est rapidement devenue dominante au point de constituer un paradigme et de donner naissance au scénario dit de spéciation par « vicariance ». Ce scénario suppose que la distribution spatiale actuelle des organismes souterrains traduit essentiellement la distribution des ancêtres de surface dont ils sont issus à partir d'un processus de spéciation allopatrique (« *Climatic relict hypothesis* » ; Barr & Holsinger, 1985) voire plus rarement d'un processus de spéciation parapatrique (« Ad*aptative shift hypothesis* », Howarth, 1987). Les organismes ayant ainsi colonisé le milieu souterrain seraient « piégés » et ne pourraient plus disperser (Lefébure, 2005). La découverte au milieu du XX^{ème} siècle d'organismes souterrains plus largement répartis notamment dans la zone hyporhéique des rivières (i.e. les sédiments saturés en eau situés dessous et le long des rivières) a réhabilité la dispersion comme un processus non négligeable façonnant les patrons de diversité (Henry, 1976 Ward & Plamer, 1994 ; Stoch, 1995 ; Holsinger, 2005 ; Culver *et al.* 2009).

Le débat entre vicariance et dispersion a longtemps monopolisé l'attention des biogéographes, alors que le rôle des facteurs environnementaux affectant la répartition de la biodiversité souterraine a été oublié (Porter, 2007). Ce débat a eu deux conséquences majeures: il a conduit à découpler la biogéographie de l'écologie et a donné un poids considérable aux facteurs historiques et à la fragmentation du milieu sans que d'autres hypothèses aient pu être testées. De ce fait, l'écologie souterraine a très peu contribué à l'avancement des connaissances en macro-écologie.

Le complexe de l'échantillonnage

L'effort d'échantillonnage a longtemps été considéré comme trop faible pour que les patrons de biodiversité aient un sens. Cette méfiance s'est traduite par une abondante littérature qui, lors de la publication des premiers patrons de diversité, s'est concentrée sur leur robustesse (Culver *et al.* 2004 ; Castellarini *et al.* 2007a ; Zagmajster *et al.* 2008, 2010 ; Dole-Olivier *et al.* 2009a ; Gibert *et al.* 2009). Culver et al. (2004) ont pourtant montré que les patrons de diversité en milieu souterrain restaient inchangés malgré la description incessante de nouvelles espèces.

Emphase sur l'histoire et le rôle de l'extinction lors des glaciations du Pléistocène

La faible taille des aires de répartition conjuguée à l'impossibilité présumée pour les organismes de disperser ont véhiculé l'idée que la faune souterraine était particulièrement sensible aux évènements paléo-climatiques de grandes ampleurs telles que les glaciations et les transgressions marines (Boutin, 1994 ; Coineau, 1994). Par conséquent, la décroissance de la richesse avec la latitude a été presque systématiquement interprétée comme une empreinte des climats froids du Pléistocène (Hof *et al.* 2008 ; Martin *et al.* 2009 ; Galassi et Stoch, 2010). Ceux-ci auraient entrainé une extinction massive dans les régions les plus septentrionales. Ces dernières n'auraient pas pu être (re)colonisées en raison des faibles capacités de dispersion des organismes (Gibert & Culver, 2005 ; Hof *et al.* 2008 ; Gibert *et al.* 2009 ; Martin *et al.* 2009).

Vers une vision plus complexe

Il a fallu attendre les années 90 pour que Stoch (1995) replace la compréhension des patrons de biodiversité souterraine dans un cadre plus large en évoquant l'influence conjointe de facteurs environnementaux tels que l'hétérogénéité de l'habitat, la surface de celui-ci, la stabilité climatique, la productivité et la quantité de ressources dans le milieu. Ainsi, l'une des premières études testant explicitement les relations entre les facteurs environnementaux et la richesse spécifique de la faune cavernicole aux Etats-Unis a apporté davantage de poids au rôle de la disponibilité en habitats plutôt qu'à celui des glaciations du Pléistocène (Christamn & Culver, 2001). En Europe, alors que des patrons de diversité plus complexe ont été mis en évidence (Culver *et al.* 2006 ; Michel *et al.* 2009), les processus et les facteurs qui les sous-tendent ont seulement été étudiés à l'échelle régionale. Dans le Jura (France), le type d'habitat et la distance aux glaciers du dernier maximum glaciaire (20.000 ans) apparaissent comme les facteurs prépondérants (Castellarini *et al.* 2007b; Dole-Olivier *et al.* 2009b). Malard *et al.* (2009) suggèrent que des facteurs régionaux seraient les plus susceptibles d'expliquer les différences de richesse entre régions européennes qui présentent des habitats semblables, mais sans en préciser la teneur. Culver *et al.* (2006) suggèrent que la crête de richesse pourrait correspondre à une zone où l'énergie productive est restée stable au cours du temps alors que les faunes du nord et du sud de l'Europe auraient subi les vicissitudes de climats historiques respectivement trop froids et trop arides. Toutefois, cette hypothèse n'a jamais été testée.

1.4) **Objectifs de la thèse**

L'objectif général de cette thèse est d'identifier et de quantifier l'influence relative des facteurs environnementaux et des processus impliqués dans la distribution spatiale de la biodiversité des eaux souterraines continentales européennes.

Tenter de répondre à cet objectif résolument ambitieux a pour contrepartie de recourir à l'utilisation de plusieurs champs disciplinaires disposant de potentiel de généralisation très variable. Deux stratégies sont alors envisageables : partir d'une inférence précise menée sur des cas particuliers pour ensuite gagner en généralité ou inversement. Dans ce manuscrit, j'ai délibérément suivi la seconde stratégie. Je pars d'un cadre général macro-écologique entrepris sur un important cortège d'espèces puis, au risque de perdre en généralité, je me concentre sur un nombre d'espèces plus restreint pour gagner en précision sur les relations entre facteurs et processus.

La déclinaison de l'objectif général en trois objectifs distincts présentés ci-dessous suit logiquement cette même stratégie.

Objectif 1 : Identifier et quantifier l'influence relative des facteurs environnementaux impliqués dans le façonnement de la biodiversité d'un milieu ayant une variabilité thermique saisonnière très réduite.

Plus précisément, il s'agit de répondre à la question, suivante : quels sont, en l'absence de saisonnalité thermique prononcée, les facteurs environnementaux actuels et historiques qui façonnent les patrons de richesse spécifique et de taille des aires de répartition des espèces en Europe ?

Pour répondre à cette question, deux grandes hypothèses sont testées :

Hypothèse 1 : La variabilité climatique à long terme, notamment au cours du Pléistocène, est un facteur clé dans la genèse du patron de taille des aires de répartition.

Hypothèse 2 : Le patron de richesse spécifique n'est pas seulement façonné par le processus d'extinction au nord causé par les oscillations climatiques du Pléistocène, mais également par la quantité de ressources trophiques disponibles et l'hétérogénéité spatiale.

Objectif 2 : Evaluer le rôle de la dispersion et plus spécifiquement celui des colonisations postglaciaires, sur les patrons d'aires de répartition.

Plus précisément il s'agit de répondre à la question suivante ; quel rôle a joué la dispersion, plus spécifiquement lors des périodes postglaciaires, sur la dynamique des aires de répartition de plusieurs espèces d'isopodes largement distribuées ?

Hypothèse : Les grandes aires de répartition traduisent une dynamique de colonisation récente (Pléistocène) qui est intervenue suite à l'ouverture de fenêtres temporelles favorables lors des phases de retrait glaciaire.

Objectif 3 : Evaluer le rôle conjoint des processus de dispersion et de sélection dans l'établissement de l'aire de répartition d'une espèce soumise aux oscillations glaciaires du Pléistocène.

Plus précisément il s'agit de répondre à la question suivante : comment interagissent les processus de sélection vis-à-vis de la température et de dispersion au cours du temps, dans l'établissement de l'aire de distribution d'une espèce confrontée à une absence de saisonnalité thermique mais subissant les oscillations climatiques du Pléistocène ?

Hypothèse : La tolérance thermique de l'espèce aurait favorisée une colonisation postglaciaire à la suite de la dernière glaciation.

Chapitre 2) Liens entre patrons de biodiversité et facteurs environnementaux par une approche macro-écologique

Les patrons de biodiversité à la surface de la planète ont très rapidement suscité un grand intérêt de la part de la communauté scientifique (Darwin, 1862 ; Wallace, 1878 ; Pianka, 1966). Historiquement, la congruence des patrons parmi des groupes taxonomiques évolutivement très éloignés a encouragé la quête d'un unique facteur universel (Davies et al. 2011). Paradoxalement, cette quête a abouti à la formulation de multiples hypothèses impliquant de nombreux facteurs (Palmer, 1994). Au cours de ces dernières années, les études de macro-écologie ont suggéré que la multi-causalité représentait certainement l'explication la plus vraisemblable (Whitakker *et al.* 2001 ; Gouveia *et al.* 2013). Ainsi, au lieu d'évaluer le support en faveur d'un seul, il s'agit désormais de quantifier l'influence relative de multiples facteurs actuels et historiques (cf. partie vers une synthèse des facteurs environnementaux).

Au cours de ce chapitre, j'emprunte cette démarche afin d'analyser, sous l'angle macro-écologique, l'influence relative des différents facteurs environnementaux (histoire, énergie et hétérogénéité spatiale) sur la taille des aires de répartition et la richesse et des crustacés aquatiques souterrains en Europe.

Répondre à cet objectif a nécessité non seulement d'acquérir une base de données d'occurrence des espèces de crustacés en Europe mais également de rassembler des données spatialisées pour différents facteurs environnementaux. L'essentiel des variables environnementales a pu être assemblé à partir de bases de données ou de cartes préexistantes. Toutefois, aucune donnée satisfaisante ne permettait de quantifier l'hétérogénéité spatiale des habitats souterrains à l'échelle européenne. L'obtention d'une telle donnée nécessitait au préalable la réalisation d'une carte typologique des habitats souterrains.

Le premier article de ce chapitre présente la première carte des habitats aquatiques souterrains à l'échelle européenne, travail auquel j'ai largement contribué à travers l'acquisition de données, l'analyse des résultats et la rédaction des résultats.

Le second article de ce chapitre présente la base de données d'occurrence des espèces de crustacés en Europe (European Groundwater Crustacean Database, EGCD) et les patrons géographiques de richesse, de la taille des aires de répartition et de la béta diversité. Il teste l'influence relative de différents facteurs environnementaux sur le patron de taille des aires de répartition. Dans ce travail, j'ai largement contribué à l'acquisition et la mise en forme des données d'occurrence, à la réalisation de l'ensemble des analyses statistiques et à la rédaction du manuscrit.

Le troisième article de ce chapitre utilise la base de données d'occurrence présentée dans l'article 2 et des données environnementales, notamment celles issues de l'article 1, afin de quantifier l'influence relative de l'histoire, de l'énergie et de l'hétérogénéité spatiale sur le patron de richesse spécifique.

2.1) Article 1 : The distribution of groundwater habitats in Europe

Jean-François Cornu, David Eme & Florian Malard.

J.-F. Cornu UMR BOREA, Département Milieux et Peuplements Aquatiques, MNHN, CNRS 7208, IRD 207, UPMC, Muséum National d ' Histoire Naturelle, 43, rue Cuvier, 75231 Paris, France.

D. Eme : F. Malard, UMR CNRS 5023, Ecologie des Hydrosystèmes Naturels et Anthropisés; Ecologie, Evolution, Ecosystèmes Souterrains, Université Lyon 1, Bât. Forel, 43, Bd du 11 Novembre 1918, 69622 Villeurbanne Cedex, France.

Article publié dans *Hydrogeology Journal*, 21: 949-960. Il a reçu une mention spéciale *Editor's Choice 2013* (Voss et al. 2014).

Abstract

Globalization and planetary environmental changes have stimulated the inventory of groundwater resources and biodiversity at continental and global scales but there has been no concurrent attempt to map the distribution of groundwater habitats even at continental scale. A vector version of the areal information contained in the international hydrogeological map of Europe (IHME) was produced, and thematic indicators for assessing its accuracy were established. Then, groundwater flow type, permeability and pore size were extracted from the vector IHME to de fi ne and map the distribution of 13 habitat types. The habitat map was used to test for latitudinal variations in habitat diversity (HD) and whether these variations might in part account for the latitudinal gradient of regional species richness. The HD of river catchments decreased significantly with increasing latitude after correcting for the effect of catchment area. HD decreased by half the amount of deviance attributed to latitude in a regression model of regional species richness, although the explanatory power of HD was probably limited by the coarse resolution of biogeographical regions. The groundwater habitat map of Europe represents a major step for the understanding, assessment and conservation of groundwater biodiversity and for incorporating ecological perspectives in groundwater management policy.

Keywords Europe, Habitat heterogeneity, Groundwater biodiversity, Geographic information systems, General hydrogeology.

Introduction

Globalization and planetary-scale changes in climate, environment and biodiversity have led science to shift from local to global perspectives. This shift has resulted in an ever-increasing generation of large-scale thematic data sets for biodiversity (GBIF 2001; IUCN 2012), climate (New et al. 2002; Tabor and Williams 2010), elevation (Farr and Kobrick 2000; Danielson and Gesch 2011), stream and catchment networks (Vorosmarty et al. 2000; Lehner et al. 2008), landcover (Loveland et al. 2000; Bartholome and Belward 2005), soil (Grunwald et al. 2011; FAO et al. 2012), geology (Durr et al. 2005; Gleeson et al. 2011) and human activity (Gaffin et al. 2004; Monfreda et al. 2008), that can be used to document and predict global changes under alternative resource use scenarios. The importance of groundwater for sustainable development and the problem of water scarcity due to water abstraction and pollution have strengthened the need for collating information on groundwater resources at continental to planetary scales (Foster and Chilton 2003; Bovolo et al. 2009; Hiscock 2011; Jones 2011; Richts et al. 2011). Indeed, one third of the world population relies upon groundwater supply for drinking water, and food production in many agricultural regions depends on the availability of groundwater for irrigation (Morris et al. 2003). Groundwater discharge also sustains the ecological function and biodiversity of many freshwater ecosystems including springs, wetlands, lakes and rivers (Hancock et al. 2005, 2009; Griebler et al. 2010; Bertrand et al. 2012). Aquifers should themselves be viewed as ecosystems because they harbor a variety of living forms among which are many groundwater obligate invertebrates with adaptive strategies for life in a dark and energy- limited environment (Danielopol et al. 2000, 2003). A total of 1,174 groundwater obligate species of crustaceans were inventoried in Europe, thereby representing more than 50 % of the number of crustacean species known from freshwater habitats in this continent (Stoch and Galassi 2010).

Hydrogeological maps and geographical information systems (GIS) depicting aquifer properties at scales ranging from continental to global are primarily intended to foster the planning, protection and monitoring of groundwater resources (Gogu et al. 2001; Stassberg et al. 2007; Struckmeier 2008). They also provide key information for inferring the distribution of distinct groundwater habitats because most recent habitat classification schemes are based upon a number of hydrogeological features including groundwater flow type, permeability, pore size and hydrological exchange with surface water (Dole-Olivier et al. 2009a; Hahn 2009). The growing agreement for a hydrogeologically based classification of groundwater habitat is one step forward for incorporating ecological perspectives in groundwater management policy (Danielopol et al. 2004; 2008; Hahn 2009; Griebler et al. 2010; Larned 2012), but there has been no attempt to build up a comprehensive map of groundwater habitats at the European scale. Yet, such a map would represent a major advance for the understanding, assessment and conservation of groundwater biodiversity in Europe (Boulton 2009; Steube et al. 2009; Larned 2012). More particularly, it would provide baseline data for testing the role of habitat heterogeneity in shaping patterns of groundwater species richness at the European scale. In the absence of quantitative criteria to assess habitat heterogeneity, latitudinal variation in subterranean species richness in Europe has so far been essentially related to differences in climate history among regions (Culver et al. 2006; Hof et al. 2008; Stoch and Galassi 2010).

Over the last 15 years, five hydrogeological thematic maps covering the European continent or significant parts of it were published at scales ranging from 1:500,000 to 1:50,000,000 (Gilbrich 2000; Hollis et al. 2002; IGRAC 2005; Wendland et al. 2008; Richts et al. 2011). Among them, the international hydrogeological map of Europe (IHME; scale: 1:500,000) is the outcome of a 50-year long international project held by the International Association of Hydrogeologists (IAH) that undoubtedly represents the most comprehensive source of hydrogeological information at the European scale (Gilbrich 2000). Although the IHME primarily aims to portray the spatial distribution

of groundwater resources in Europe, the hydrogeological and lithological features of aquifer types are sufficiently detailed to be used as a basis for a classification and mapping of distinct groundwater habitats in Europe. Yet, the digital image format of the IHME has restricted its use in GIS, which often forms the basis of continental-scale projects in groundwater ecology and hydrogeology (Deharveng et al. 2009; Wodja et al. 2010; Richts et al. 2011; MacDonald et al. 2012). Although the digitalization of geological and hydrogeological paper maps was initiated in the 1990s by a number of institutional providers (Laxton and Becken 1996), Gilbrich (2000) pointed out that the digitalizing of the IHME "was not foreseen at this stage in view of the enormous costs which such an exercise would incur". Since then, Nikas et al. (2010) provided a GIS of the Athina IHME sheet covering Greece; however, a vector version of the whole IHME is still lacking.

The objective of the present study was to derive a comprehensive map of groundwater habitats in Europe from the hydrogeological information contained in the IHME. First, a vector version of the IHME was produced by digitalizing as polygons the areal information representing hydrogeological and lithological features of the rocks. Second, the distribution of groundwater habitats was classified and mapped using a limited set of key biologically relevant variables that were extracted from the vector IHME. Third, latitudinal variation in groundwater habitat diversity was determined, followed by an assessment of whether this diversity might in part account for the observed latitudinal gradient of groundwater biodiversity in Europe.

Materials and methods

Description of the IHME

For the sake of clarity, a short and simplified description of the IHME is provided (see Struckmeier and Margat (1995) and Gilbrich (2000) for a detailed description). The IHME is a series of 25 paper sheets at scale 1:1 500 000 realized between 1970 (C5 sheet, Bern) and 2008 (D6 sheet, Athina) under the auspices of IAH and UNESCO. They were recently georeferenced by BGR and delivered in digital format (img format) with their accompanying detailed legend (BGR 2012, see figure in electronic supplementary material 1 (ESM1). Hydrogeological features of the rocks are represented with colors using a two-level classification. The first level, referred to as groundwater flow type, distinguishes between three rock categories: (1) porous rocks in which flow is mainly intergranular (colored blue); (2) fissured rocks, including karstified rocks (colored green) and; (3) rocks with little or no groundwater flow (colored brown). At the second level, referred to as productivity, porous and fissured rocks are both subdivided into two categories represented with distinct color tones: (1) highly productive aquifers (dark color tone) and; (2) low and moderately productive aquifers (light color tone). Similarly, rocks containing limited groundwater resources and rocks with essentially no groundwater are shown with light and dark browns, respectively. In the legends accompanying the 25 georeferenced IHME sheets, a range of permeability is provided for each of the six flow-type productivity combinations de fi ned in the preceding. The lithology of strata at outcrop is shown with grey ornaments beneath the colors (not visible in figure in ESM1, but see Fig. 2). For each ornament represented on the map, the lithological composition of strata is further detailed in the legend and its approximate age is provided by means of stratigraphic symbols. A number of detailed hydrogeological, hydrological and geological information (e.g. springs, streams, and faults) are further represented with symbols and lines.

Vectorization and attribution of areal information

Successive steps in the generation of the vector version of the IHME and groundwater habitat map of Europe are described in Fig. 1. The first step of the digitalization of the IHME consisted of vectorizing as polygons the arealinformation representing hydrogeological and lithological features of the rocks (Masuch-Oesterreich 2000; Fig. 2). A polygon was de fi ned as a continuous area corresponding to a single legend and a single flow-type productivity category (color tones). Polygon vectorization was done manually because the complexity of ornaments and the resemblance of color tones among map sheets prevented the use of an automatic or semi-automatic procedure. A working scale of 1:200,000 was retained as a reasonable choice to preserve the accuracy of polygon contours and to keep as many small polygons as possible in the final digital database without inflating excessively the burden of the vectorization task. A set of 25 polygon shape files covering all IHME sheets were created, merged and converted into a single ArcINFO coverage. Topological tools were used to clean silver polygons, overlaps, and gaps prior to exporting the coverage as feature class into a personal geodatabase. Vectorization of the 25 IHME sheets was completed within 1 year under ArcGIS 9.3 software (ESRI 2010).

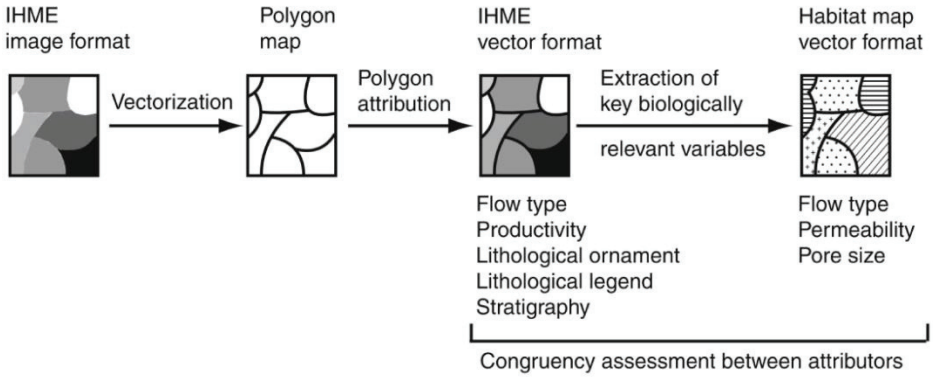


Figure 1: Successive steps in the generation of the vector version of the IHME and groundwater habitat map of Europe.

The second step of the digitalization consisted in attributing a set of hydrogeological and geological features to each polygon previously vectorized. Each feature was recorded in a distinct field of the feature class attribute table. The first two fields — successively entitled flow type and productivity — correspond to the two levels of the classification scheme used to describe hydrogeological features of the rock. The next three fields contain lithological information derived from the ornaments and their associated legends. The field “lithology_ornament” employs a codification system to provide a description of the lithology as indicated by the ornaments. The 25 sheets of the IHME collectively contain a total of 33 simple ornaments describing general lithological types (e.g. limestone, sandstone) and 98 combined ornaments describing strata of varying lithology (e.g. limestone and sandstone). The field “lithology_legend” contains the detailed lithological description of strata as provided in the map sheet legends for each ornament. The IHME contains a total of 1,105 lithological legends. Finally, the field “stratigraphy” indicates the period and epoch of lithological strata using the denominations of the International Geological Map of Europe (scale: 1: 1 500 000) as references.

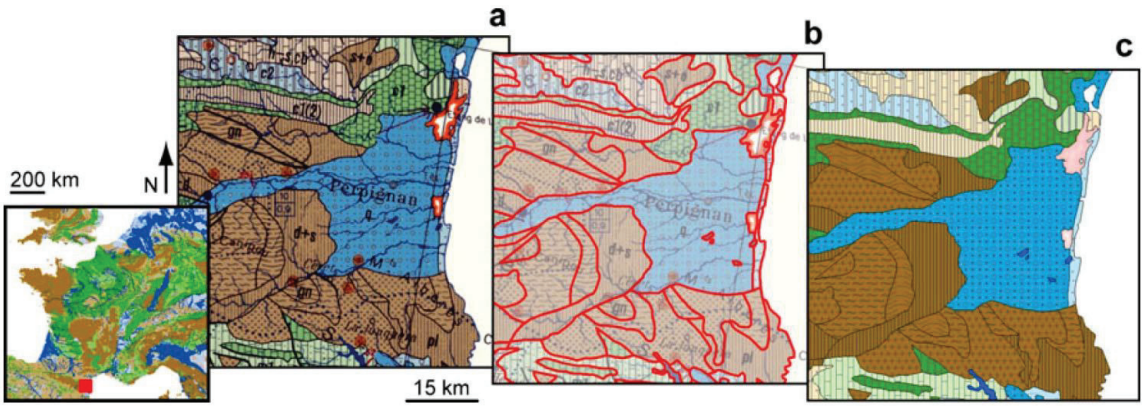


Figure 2: Digitalization of the IMHE. **a** Inset of the IHME (image format); **b** vectorization as polygons of the areal information; **c** attribution of hydrogeological and geological features to polygons.

Semantic interpretation and congruency assessment

For the 25 sheets of the IHME, the attribution of features to polygons was performed independently by two operators to assess the fidelity of the vector map information with that of the image map. The incongruence between the semantic interpretations conducted by two operators only provides a likelihood correctness indicator because none of the two individual interpretations can be considered as a reference (Chrisman and Lester 1991). First, the feature attribution was performed for all polygons of five map sheets (B5 London, C5 Paris Sud, C4 Berlin, C5 Bonn and D5 Budapest). For these five map sheets, crossed tabulations of semantic interpretations revealed that the occurrence of small polygons and legends with similar ornaments were the major sources of incongruence. Second, a stratified random design was used to assess the incongruence of the 20 other map sheets (Foody 2002; Stehman 2000, 2009). Three classes of polygon size (<10, 10 – 100 and >100 km^2) were distinguished and five polygons were selected randomly in each combination of lithological legend and polygon size classes. The maximum number of polygons that could be selected for a given map sheet comprising n lithological legends was 5×3×n polygons, although this upper limit was almost never met. This sampling design could overweight the representation of rare combinations because it did not take into consideration the frequency distribution of polygon size and lithological legends. However, it ensured that all combinations would be represented. Legend incongruence was expressed as the proportion of both polygons and area that was attributed to different legends. This overall quality index was selected among the many indices available for comparing categorical maps (Stehman 1999; Foody 2007; Liu et al. 2007) because it was comprehensible to all map users. Similarly, lithological and habitat incongruences were calculated as the proportion of polygons and area attributed to different lithological ornaments and groundwater habitats, respectively (see the following).

Definition and mapping of groundwater habitats

The classification of groundwater habitats was based upon three distinct criteria: (1) groundwater flow type; (2) permeability; and (3) void size. First, consolidated and unconsolidated rocks were considered separately because they provide distinct microhabitats colonized by different species assemblages (Gibert et al. 1994a; Malard et al. 2009). Fissures are the only voids accessible to the fauna in consolidated rocks, although they can be considerably enlarged by weathering in limestone rocks. Some consolidated rocks (i.e. chalk, sandstone) combine intergranular and fissure flows, but intergranular voids are too small to harbour even meiofaunal organisms (body size: 63 – 1,000 μm). The voids between grains are the primary microhabitats in unconsolidated rocks but a

number of organisms can affect the size and distribution of voids through their bioturbation activity (Datry et al. 2003). Second, three classes of permeability — high, moderate and low — were distinguished according to the information provided in the legend of the IHME. Permeability is a measure of the interconnectedness of the voids which influences both the movement of animals and the flux of nutrients, dissolved oxygen and organic carbon. Finally, void size was considered as a biologically relevant criterion because it determines the range of species having different body sizes that can co-occur in an aquifer. Lithological information was used to distinguish between large and small void sizes, although the range of void sizes is typically much smaller in unconsolidated rocks than in consolidated rocks. Gravelly sediments were considered to exhibit large void size as compared to sandy, silty and clayey sediments. Similarly, karstified rocks such as limestone, dolomite, and chalk were classified as having large void size as compared to other nonkarstified rocks. This three-criteria classification (2 groundwater flow types × 3 permeability classes × 2 void sizes) resulted in 12 habitat categories to which a 13th category representing non-aquiferous rocks was added. Habitat categories were added as a field in the attribute table of the vector IHME.

Latitudinal patterns of habitat diversity and groundwater biodiversity

Latitudinal variation in groundwater habitat diversity at the European scale was determined, followed by an assessment of whether this diversity might in part account for the decrease in groundwater biodiversity with increasing latitude. First, habitat heterogeneity (H) was calculated for a total of 778 European river catchments using the Shannon index ($H= - \Sigma\ p_i \times \ln\ p_i$), where p_i represented the areal proportion of each groundwater habitat category represented within a catchment. The selected catchments corresponded to the sea outlets of the pan European river and catchment database, having an area higher than 500 km^2 (Vogt et al. 2007, see figure in ESM2). In order to account for the effect of area on heterogeneity, the residuals of the linear relationship between habitat diversity and catchment area were regressed against the latitude of catchment centroids. The regression between the residuals and latitude was assessed using a generalized linear model with linear and quadratic terms. The reduction in deviance associated with each term was tested for significance at α=0 0.05 using a χ^2 test (Venables and Ripley 2002). The residuals of both relationships were tested for normality using a Shapiro-Wilk test.

Second, a generalized linear model was performed to examine whether variation in habitat diversity among biogeographical regions could complement the analysis of groundwater biodiversity patterns at the European scale implemented in Hof et al. (2008). These authors compared the latitudinal patterns in regional species richness (RSR) among species adapted to three distinct habitat types: groundwater, lotic (running water) and lentic (standing water) habitats. The spatial extent of their study covered Europe and the grain size corresponded to biogeographical regions de fined by Illies (1978; see figure in ESM2). Hof et al. (2008) used multiple linear regressions to test for the effect of latitude on RSR after accounting for the effect of area, longitude, and elevation of regions (see table in ESM1). Differences in the latitudinal variation of RSR among species adapted to the three habitat types were attributed to differences in the propensity for dispersal. Particularly, the monotonic decline of groundwater and lotic species richness with increasing latitude was attributed to a low recolonization rate of northern regions affected by Pleistocene glaciations. Yet, the heterogeneity within the three habitat types across regions was not considered in the analysis, although Hof et al. (2008) acknowledged that it might also explain differences in the latitudinal variation of RSR. Here, statistical analysis was performed to test whether the effect of latitude on groundwater species richness was still significant after accounting for the effect of habitat diversity, which was calculated for each region using the Shannon index (see preceding formula). Area, longitude, elevation, habitat diversity and latitude of regions were included as quantitative variable in a generalized linear model and the

reduction in deviance associated with each variable was tested for significance at α0 0.05 using a χ^2 test (Venables and Ripley 2002). Statistical analyses were performed in R (R Development Core Team 2006).

Results

The vector IHME

The vector IHME comprised 61,275 polygons and the five thematic fields of its attribute table enabled the quantification of the surface area of distinct categories of groundwater flow type, aquifer productivity, and lithology in Europe (see figure in ESM1, table in ESM2 and table in ESM3). Highly productive aquifers occupied only 19.5 % of the European territory among which 56 and 44 % were represented by fissured and porous rocks, respectively. Yet, limestone rock formations comprising aquifers of usually high vulnerability to human activities represented 79.4 % of the surface area of highly productive aquifers in fissured rocks, although the latter occurred in not less than 26 distinct lithologies.

Congruency assessment of the vector IHME

The average proportion of polygons attributed to different legends by two distinct operators was 12.9 ±2.9 and 9.5 ±4.3 % for the fully sampled sheets (n = 5) and subsampled sheets (n 0 20), respectively (Table 1). Three sheets among subsampled sheets had distinctively higher incongruence values (sheets D4, D6 and E5). Yet, legend incongruence expressed as an area proportion decreased to 4.2 ±1.9 and 4.8 ±4.7 % for the fully sampled and subsampled sheets, respectively. The high incongruence in subsampled sheet E3 (i.e. 22.4 % for area) was due to the differential attribution of a single large polygon, the ornament of which corresponded to two distinct legends. Decrease in legend incongruence when expressed as an area proportion reflected the negative relationships between incongruence and polygon size (Fig. 3). Indeed, the average legend incongruence for small (<10 km^2), medium (10 – 100 km^2) and large (>100 km^2) size polygons was 18.0 ±7.7, 8.2 ±4.9 and 4.6 ±2.7 %, respectively (Table 1). Incongruence became successively smaller when calculated as the average proportion of polygons attributed to different legends (10.2 ±4.2 %), lithological ornaments (7.9 ±3.9 %), and groundwater habitats (3.7 ±2.4 %; Fig. 3). Lithological and habitat incongruence were also smaller when expressed as a proportion of area rather than polygons. The average areal proportion attributed to different habitats by two operators was only 0.9 ±0.9 %.

Table 1: Congruency assessment of the vector version of the international hydrogeological map of Europe.

Map sheet (see Figure ESM1)		Sampling [a]		Legend incongruence [b] Total		Polygon size classes			Ornament [c] incongruence		Habitat [d] incongruence	
Code	No. of polygons	Polygons (%)	Area (%)	Polygons (%)	Area (%)	< 10 km^2	10-100 km^2	>100 km^2	Polygons (%)	Area (%)	Polygons (%)	Area (%)
A5	964	40.1	35.8	7.8	1.8	13.9	9.5	2.3	7.6	1.8	3.6	0.6
A6	1102	42.9	30.7	8.9	2.4	18.7	8.8	3.6	7.5	1.9	5.2	1.0
B2	1184	28.2	46.0	11.5	6.4	22.0	9.9	4.9	9.3	0.5	9.3	0.5
B3	1670	26.2	34.4	12.1	5.2	29.1	3.4	3.9	9.7	5.1	2.1	0.8
B4	2195	100.0	100.0	10.7	1.0	12.5	1.4	0.0	8.1	0.6	3.3	0.3
B5	3134	100.0	100.0	15.3	5.7	17.7	8.3	6.6	8.0	2.4	3.8	1.5
B6	2058	29.4	29.2	11.6	3.6	17.1	14.1	4.2	4.7	0.9	3.9	0.8
C2	1612	22.0	29.9	6.8	2.2	14.0	3.2	3.2	6.0	2.0	1.1	0.0
C3	4577	19.7	24.5	8.7	2.4	19.0	4.5	3.4	7.5	1.6	4.1	0.3
C4	2793	100.0	100.0	13.8	4.5	16.5	8.4	5.4	13.3	4.2	8.1	3.0
C5	4827	100.0	100.0	15.6	4.6	18.3	5.2	3.0	10.0	2.1	6.8	1.1
C6	1030	41.1	38.6	10.9	3.4	20.2	10.7	5.0	8.6	2.6	3.1	0.5
D2	6022	10.8	10.2	5.4	0.6	11.5	3.5	1.5	5.4	0.6	1.5	0.0
D3	5388	9.5	51.5	5.5	7.1	9.7	6.4	2.3	4.4	3.4	1.6	3.3
D4	2755	46.2	31.4	14.6	4.2	37.2	15.8	10.0	12.7	2.9	5.6	1.3
D5	3219	100.0	100.0	8.9	5.0	10.2	5.5	5.7	7.3	4.1	2.7	1.3
D6	3500	22.9	26.0	16.5	8.6	26.5	13.7	9.9	14.9	7.3	5.6	2.0
E2	2006	37.8	46.8	5.9	5.2	11.8	5.1	4.0	2.1	0.6	1.5	0.0
E3	1134	47.4	41.6	6.8	22.4	13.6	5.5	6.7	4.4	4.0	1.4	0.0
E4	1024	41.6	25.2	6.6	1.3	17.3	7.2	2.1	5.9	1.2	2.9	0.5
E5	1070	64.2	57.6	22.1	7.3	39.2	24.2	11.7	19.8	5.3	8.4	1.7
E6	2827	27.8	23.4	7.0	3.6	16.7	5.9	2.7	5.8	3.5	1.6	0.6
F2	2286	29.3	20.0	7.0	3.3	13.6	5.4	5.1	6.6	3.2	2.1	0.4
F3	1091	34.3	27.5	7.9	3.2	14.3	10.9	5.0	4.9	1.9	0.8	0.0
F4	1807	26.2	19.9	6.7	1.9	10.2	9.4	3.0	4.1	1.0	1.8	0.7

[a] Proportion of map sheet expressed as a percentage of the number of polygons and area representing rocks (water bodies and glaciers are excluded).

[b] Incongruence expressed as the proportions of polygons and area that was attributed to different legends,

[c] Incongruence expressed as the proportions of polygons and area that was attributed to different ornaments,

[d] Incongruence expressed as the proportions of polygons and area that was attributed to different habitats.

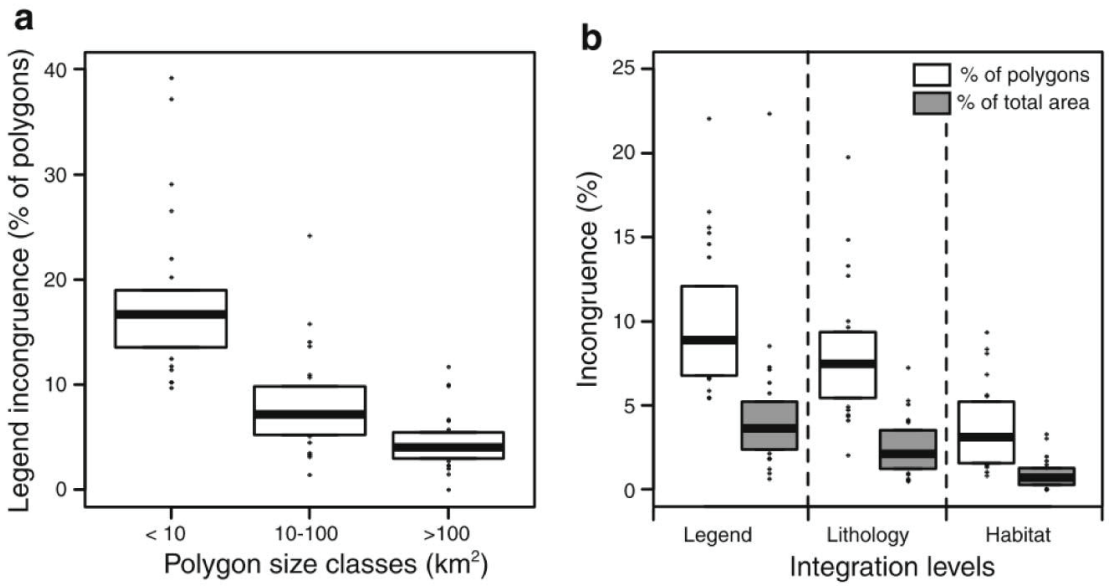


Figure 3: Congruency assessment of the vector IHME. **a** Differences in legend incongruence among three polygon size classes. **b** Differences in incongruence, expressed as a proportion of polygons and area, among three integration levels. The horizontal bar, boxes and black dots show the median, interquartile distance and outliers, respectively (n=25 sheet maps)

The groundwater habitat map of Europe

Several general patterns emerged from the mapping of groundwater habitats in Europe (Fig. 4). First, unfavorable habitats corresponding to practically non-aquiferous rocks and aquifers of low permeability with small pore size occupied 43.7 % of the total land area. This proportion reached more than 85 % in the major part of Scandinavia including Norway, Sweden and Finland. Second, there was a marked difference in the spatial con fi guration of habitats between southern and mid Europe. Habitats exhibited a patchy distribution in southern peninsulas including the Iberian, Balkanic and Turkish regions, whereas they expanded over large and continuous areas at mid latitudes. More particularly, there was a continuous ridge of favorable habitats in unconsolidated sediments of moderate to high permeability with large pore size that expanded over a distance of 2,000 km along the southern shores of the North and Baltic seas. Habitat diversity increased significantly with increasing catchment area ($r=0.24$, $p<0.001$, Fig. 5). The residuals of the relationship between habitat diversity and catchment area were not normally distributed (Shapiro-Wilk test; $p<0.001$; $W=0.9836$) and decreased significantly with increasing latitude of catchment centroids (Fig. 5; see also figure in ESM2). The effect of latitude was best fitted using a second-order polynomial function (likelihood ratio test; $p<0.001$). The residuals of the relationship between habitat diversity and latitude after removing for the effect of area were not normally distributed (Shapiro Shapiro-Wilk test; $p<0.001$; $W=0.992$).

Effect of habitat diversity on species richness pattern

Maximum elevation and habitat diversity produced a significant reduction in deviance in the regression model of regional groundwater species richness (Table 2). Habitat diversity accounted for 25.6 % of model deviance. Yet, latitude was still significant and represented 31.3 % of model deviance even after accounting for the effect of habitat diversity. Habitat diversity decreased significantly with

increasing latitude (r =− 0.48, p=0.02); however, this negative correlation was driven by the low habitat diversity of two northern latitude regions (the Borealic uplands and Fenno-scandian shield) (Fig. 6). Removing these two outliers resulted in a nonsignificant correlation between habitat diversity and latitude (r=− 0.16, p=0.502), whereas the negative correlation between latitude and groundwater species richness remained significant (r =− 0.76, p<0.001; Fig. 6). Also, the deviance accounted for by habitat diversity in the regression model of regional groundwater species richness was no longer significant (deviance 2.7 %, p 0 0.37) when the Borealic uplands and Fenno-scandian shield were no longer considered in the analysis.

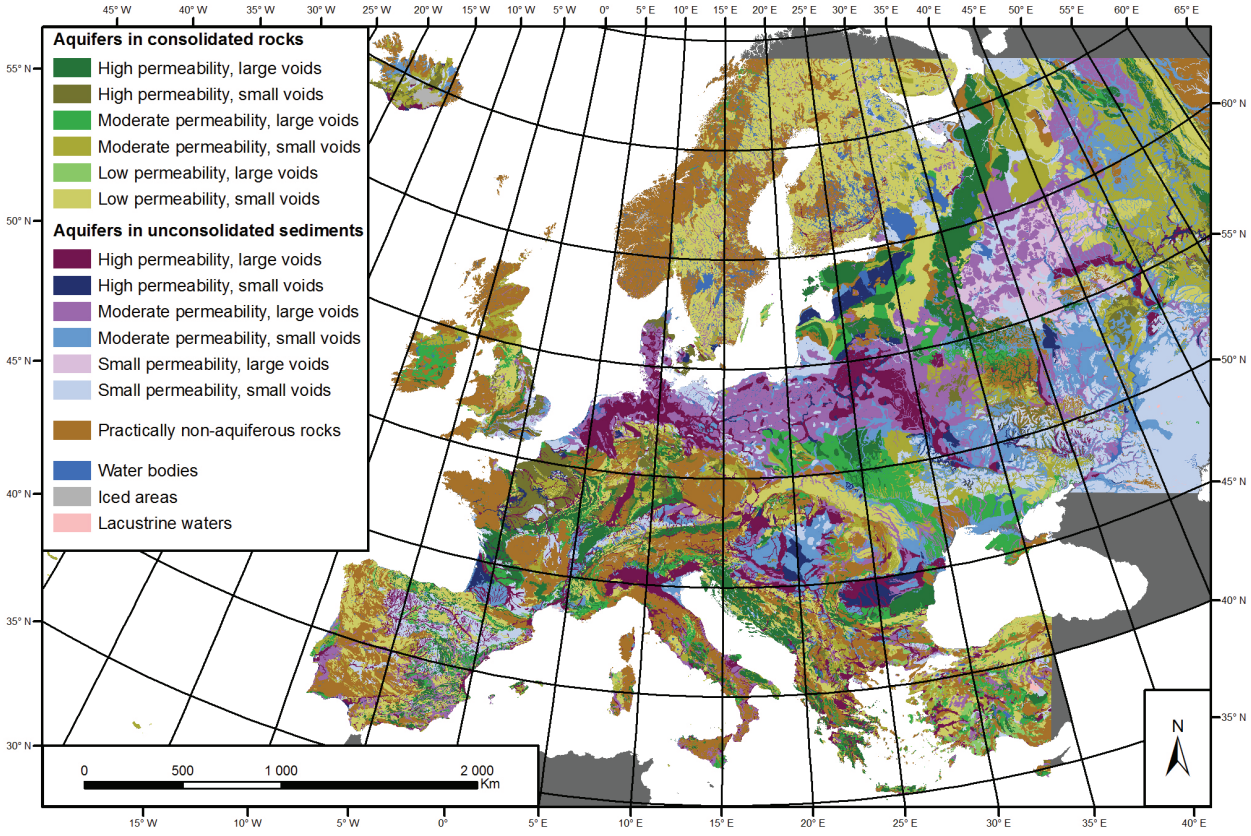


Figure 4: The groundwater habitat map of Europe (Lambert azimuthal equal-area projection).

Discussion

From aquifer types to groundwater habitats

The objective of this study was to map the distribution of groundwater habitats in Europe using information contained in the IHME. To do so, much effort was put in developing a GIS that gathered, in an explicit vector format, the areal information contained in the 25 sheets of the IHME. In the conceptual framework proposed by Gogu et al. (2001) for the development of GIS in hydrogeology, this ready-to-use dataset represents the first European vector "aquifer system geometry". This GIS version can easily be updated to accommodate additional coverages containing detailed geological and hydrogeological data represented in the IHME such as faults, springs, groundwater quality data and human alterations of the natural groundwater regime (Struckmeier and Margat 1995). Efforts made to digitalize the IHME are small compared to the 50-year-long

collaborative effort needed to achieve the original paper sheets (Gilbrich 2000; Struckmeier and Margat 1995); however, this vector version may further a wider use of the IHME in future projects dealing with the multiple functions of groundwater. In that sense, the groundwater habitat map which focuses on the ecological function of aquifers can be seen as the first of many possible derivatives of the vector version of the IHME (Struckmeier 2008).

The production of the vector IHME and its use in an ecological context required solving two major difficulties. First, the risk of misrepresenting the original hydrogeological information during the digitalizing process had to be evaluated. Second, classifying and mapping groundwater habitats required that a limited set of key biologically relevant variables could be extracted from a hydrogeological map that originally focused on the concept of aquifer productivity. On average, the incongruence between semantic interpretations made by different operators was 10.2 % when expressed as a proportion of polygons, indicating that hydrogeological information contained in the image map sheets was accurately reproduced in the vector map. The sampling design which represented equally all combinations of lithological legends and polygon size classes slightly underestimated inconsistency in subsampled sheets because differences in interpretation between operators was typically higher for small size polygons. Yet, the uneven distribution of misinterpretations among size classes resulted in a much lower average inconsistency when expressed as an area proportion (i.e. 4.7 %). This inconsistency even became smaller for higher integration features including lithology (2.6 %), habitats (0.9 %) and aquifer productivity categories (< 1 %, data not shown). This ensures the use of the vector map in future groundwater projects because quantitative estimates that can be derived from it are unlikely to be biased by misinterpretations of image map sheets. To some extent, the strong convergence among interpretations from several operators also validates past efforts made to select symbols, ornaments and colours that are internationally recognized (IAH 1983; Struckmeier and Margat 1995). In fact, the strongest inconsistencies in map sheets D4, D6 and E5 could substantially be improved by using the original paper map sheets because they were essentially due to the poor quality of scanned image maps.

The habitat classification was based upon three hydrogeological criteria that had repeatedly been proved to influence the composition and distribution of groundwater organisms: flow type, void size and permeability (Dole-Olivier et al. 2009b; Hahn and Fuchs 2009). Attributing flow type categories (i.e. intergranular versus fissure flow) to each polygon did not present any major difficulties because these two categories almost entirely corresponded to the three categories distinguished in the first hierarchical level of the hydrogeological map (i.e. porous rocks, fissured rocks, and rocks with little or no groundwater flow). Similarly, the distinction between large and small void sizes appeared to be straightforward for the aquifers in consolidated sediments as grain-size composition was sufficiently described in the lithological description of ornaments. However, this distinction was not always practical for the aquifers in consolidated rocks because the significance of karstification of carbonated rocks was not systematically provided in all sheets of the IHME. The largest uncertainty concerned the attribution of permeability categories (i.e. high, moderate and low) to each polygon partly because the IHME was based on categorizing the concept of aquifer productivity rather than rating well yield (IAH 1983; MacDonald et al. 2005). The concept of aquifer productivity is typically more appealing to groundwater managers and decision-makers than the mere ability of a well to yield water, but it encompasses several characteristics such as permeability, extent of aquifer and its thickness. For example, the productivity category entitled “local or discontinuous productive aquifers or extensive, but only moderately productive aquifers” recovered two distinct permeability categories (high and moderate) that can hardly be distinguished without an expert knowledge of regional hydrogeology. Thus, attributing permeability categories from verbal description of aquifer

productivity turned out to be more difficult than linking permeability to quantitative estimates of well yield, although the latter may be strongly dependent upon well design efficiency.

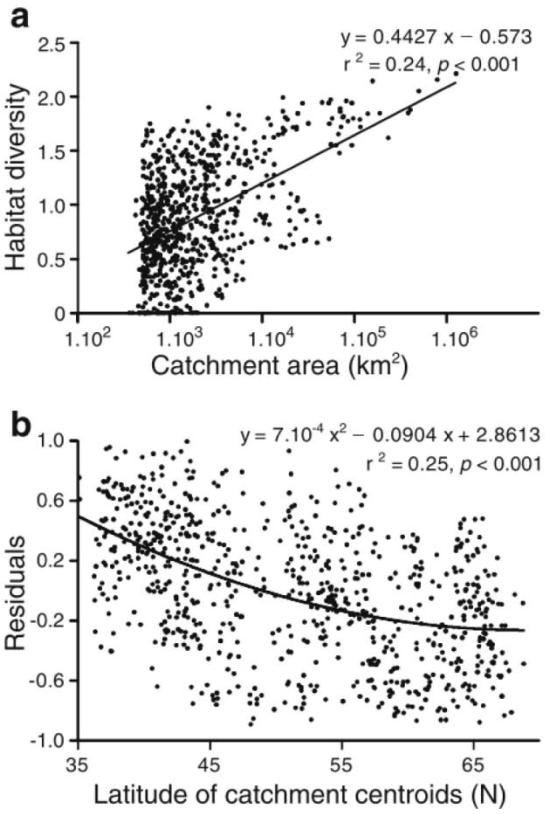


Figure 5: Latitudinal variation in habitat diversity. **a** Relationships between habitat diversity and catchment area (n=778). **b** Relationships between habitat diversity and latitude of catchment centroids after removing for the effect of catchment area (residuals).

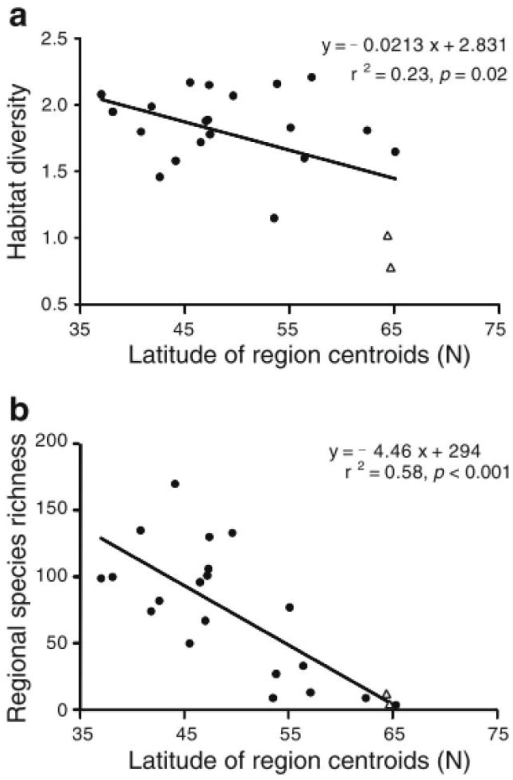


Figure 6 : Relationships between **a** habitat diversity, **b** regional groundwater species richness, and latitude of the centroids of biogeographical regions (n=22). White triangles correspond to the Borealic uplands and Fenno-scandian shield.

Linking habitat diversity and biodiversity patterns

Identifying processes shaping groundwater biodiversity at multiple spatial scales requires elaborating hierarchical models and quantifying biologically relevant variables at scales ranging from local to global. Since the 1980s, conceptual linkages between groundwater ecology and hydrogeology have stimulated synthetic approaches to investigate groundwater biodiversity patterns (Rouch 1986; Gibert et al. 1994a). A common conceptual framework consists in viewing groundwater landscapes as a series of nested spatial units including ecoregions, catchments, and aquifers with their recharge and discharge areas (Gibert et al. 1994b; Hahn 2009; Malard et al. 2009). Several studies showed that hydrogeological variables such as permeability and void size largely determined biodiversity patterns at spatial scales ranging from local to regional (Malard et al. 2002; Dole-Olivier et al. 2009b; Hahn and Fuchs 2009). The present study went one step further by providing a hydrogeologically based map of groundwater habitats that could be used to document patterns of habitat diversity and their effect on groundwater biodiversity at the European scale. Habitat patches as represented in the map had an

average area of 315±4,257 km^2 (n=29,201), which was considerably smaller than that of river catchments (10,669 ±66,221 km^2, n=778 sea outlets) but larger than the area of most aquifers in Europe.

Habitat diversity of catchments was found to decrease with increasing latitude after correcting for the effect of the catchment area, a result which was consistent with the higher compositional heterogeneity and patchiness of habitats in southern Europe. However, the first attempt to integrate habitat diversity as a quantitative variable into explanatory models of large-scale groundwater biodiversity patterns indicated that variation of habitat diversity could not account alone for the decrease in regional species richness with increasing latitude. Integrating habitat diversity into the generalized linear model decreased by half the amount of deviance attributed to latitude in the original model of Hof et al. (2008). However, latitude still produced a significant reduction in deviance. Moreover, most of the deviance attributed to habitat diversity arose from the low species richness of two northern regions, the groundwater habitats of which were over-represented by practically non-aquiferous rocks and aquifers in consolidated sediments of low permeability and small pore size. These findings reinforced the view that the present-day latitudinal pattern of regional species richness might still retain the imprint of Pleistocene glaciations (e.g. high species extinction rates in northern Europe) because of the weak propensity for dispersal among groundwater organisms (Hof et al. 2008; Stoch and Galassi 2010). However, the importance of habitat diversity in explaining the latitudinal gradient of species richness might have in part been hindered by the coarse spatial resolution of biogeographical regions. Moreover, biogeographical regions as defined by Illies (1978) may not necessarily be suited for exploring the determinants of groundwater biodiversity patterns (Stein et al. 2012). Testing for the effect of increasing spatial resolution on the explanatory power of habitat diversity would require elaborating species occurrence databases that can be used to compute species richness of smaller spatial units (e.g. river catchments and 100×100-km grid cells), an effort which is actually being made within the framework of the European Biofresh project (Malard 2012).

The usefulness of the European groundwater habitat map for analyzing biodiversity patterns could benefit from three methodological developments. First, small but potentially species-rich groundwater habitats may have escaped the spatial resolution of the IHME (i.e. 1:1,500,000). This is particularly true for the hyporheic zone of streams some of which may flow over practically non-aquiferous rocks. Mapping the extent and hydrogeological features of hyporheic habitats (e.g. permeability and sediment size) would require characterizing simultaneously key features of catchments such as the stream network, parent lithology, and the overland flow erosive capacity (Valett et al. 1996; Buffington and Tonina 2009). Second, the categorization of groundwater habitats, which is entirely based upon hydrogeological criteria, would greatly benefit from the integration of key hydrochemical parameters such as the concentration of dissolved oxygen (Malard and Hervant 1999). Indeed, low oxygen supply can severely restrict species occurrence in several aquifers recognized as potentially favorable habitats, in particular those expanded along the southern shores of the North and Baltic seas (Stein et al. 2012). The groundwater habitat map could be amended to accommodate hydrochemical data sets available across Europe for several aquifer types (Wendland et al. 2008). Third, continuous layers of habitats may in fact be highly fragmented because of the occurrence of many aquifers with well-defined boundaries. This limitation is unavoidable because the IHME does not delineate aquifers but rather shows the contours of aquifer types. Aquifer maps are

Table 2: Results of the generalized linear model for testing variables for effects on regional groundwater species richness.

	Degree of freedom.	Deviance	Residual degree of freedom	Residual deviance	Probability (p)
Null	1		21	53 334	
Area	1	2 013.5	20	51 320	0.210
Longitude	1	28.2	19	51 292	0.882
Maximum altitude	1	9 921.4	18	41 370	0.005
Habitat diversity	1	8 767.9	17	32 602	0.009
Latitude	1	10 711.7	16	21 891	0.004
Latitude ^{2}	1	2 673.7	15	19 217	0.149

becoming increasingly available at region to country scales, but the authors are unaware of any map that delineates aquifers with sufficient detail at European scale. Amending the habitat map with aquifer boundaries would enable computation of spatial indices of landscape pattern for testing the effect of habitat configuration (e.g. habitat size and fragmentation) on groundwater biodiversity (Rutledge 2003). This is a promising perspective, as the difference in spatial con figuration of habitat types between northern and southern latitude regions suggests that habitat fragmentation may in part explain the high number of groundwater endemic species in southern Europe.

Habitat diversity and the assessment and conservation of groundwater biodiversity

Despite its limitations, the groundwater habitat map should also be conceived as a useful tool for the assessment and conservation of groundwater biodiversity at the European scale. Indeed, knowledge of the distribution and spatial extent of habitats among regions is a crucial step for improving the efficiency of large-scale species inventories which necessarily aim to capture heterogeneity among habitats (Castellarini et al. 2007; Hahn and Fuchs 2009; Dole-Olivier et al. 2009a). While effort is continuously being made to improve species inventories (Deharveng et al. 2009), the use of abiotic and biotic surrogates for biological diversity are increasingly considered as cost and time-effective methods for assessing biodiversity and developing conservation strategies (Ward et al. 1999; Stoch et al. 2009). Recent applications of reserve selection methods for designing networks of groundwater-protected sites at the European scale were based solely on the achievement of species representation goals (Michel et al. 2009). Yet, the vector map of groundwater habitats may be used to integrate habitat representation goals in the design of reserve networks, thereby acknowledging uncertainties in species distribution. Finally, classifying and mapping habitats across Europe is one of the key requirements for implementing issues related to the maintenance and management of groundwater ecosystems and biodiversity in the European Union (EU) Groundwater Directive (Danielopol et al. 2008; Hahn 2009). This necessary implementation may turn easier as the growing convergence between ecology and hydrogeology, sometimes referred to as hydrogeoecology (Hancock et al. 2005, 2009; Humphreys 2009), enables the provision of operational tools that truly meet the needs of resource managers.

Acknowledgements This research was partly funded by the European Commission (7th EU Framework Programme, Contract No. 226874, BioFresh) and the Agence Nationale de la Recherche (ANR 08JCJC012001, "DEEP"). We thank Laurent Simon and Christophe Douady for commenting

on an earlier draft of this manuscript and the associate editor and two anonymous reviewers for their valuable comments. The groundwater habitat map of Europe can be downloaded from the web site of the European BioFresh project (http://data.freshwaterbiodiversity.eu/data/shape fi les/). The vector version of the IHME can be obtained upon request from the lead author. In return for the use of these two vector maps, we kindly request that you cite the present publication.

References

Bartholome E, Belward AS (2005) GLC2000: a new approach to global land cove mapping from Earth observation data. Int J Remote Sens 26(9):1959 – 1977

Bertrand G, Goldscheider N, Gobat JM, Hunkeler D (2012) From multi-scale conceptualization to a classification system for inland groundwater ecosystems. Hydrogeol J 20:5 – 25

BGR (Bundensanstalt für Geowissenschaften und Rohstoffe) (2012) IHME 1500: international hydrogeological map of Europe. Download at http://www.bgr.de/app/ fishy/ihme1500/ download.html. Accessed March 2013

Boulton AJ (2009) Recent progress in the conservation of groundwaters and their dependent ecosystems. Aquat Conserv 19:731 – 735

Bovolo CI, Parkin G, Sophocleous M (2009) Groundwater resources, climate and vulnerability. Environ Res Lett 4(3):1 – 3

Buffington JM, Tonina D (2009) Hyporheic exchange in mountain rivers II: effects of channel morphology on mechanics, scales, and rates of exchange. Geog Compass 3:1 – 25. doi:10.1111/j.1749-8198.2009.00225.x

Castellarini F, Dole-Olivier M-J, Malard F, Gibert J (2007) Using habitat heterogeneity to assess stygobiotic species richness in the French Jura region with a conservation perspective. Fund App Limn 169:69 – 78

Chrisman N, Lester M (1991) A diagnostic test for error in categorical maps. AutoCarto 10:330– 348

Culver DC, Deharveng L, Bedos A, Lewis JJ, Madden M, Reddell JR, Sket B, Trontelj P, White D (2006) The mid-latitude biodiversity ridge in terrestrial cave fauna. Ecography 29:120 –128

Danielopol DL, Pospisol P, Rouch R (2000) Biodiversity in groundwater: a large-scale view. Trends Ecol Evol 15(6):223 –224

Danielopol DL, Griebler C, Ginatilaka A, Notenboom J (2003) Present state and future prospects for groundwater ecosystems. Environ Conserv 30:104 – 130

Danielopol DL, Gibert J, Griebler C, Gunatilaka A, Hahn HJ, Messana G, Notenboom J, Sket B (2004) Incorporating ecological perspectives in European groundwater management policy. Environ Conserv 31:185 – 189

Danielopol D, Griebler C, Gunatilaka A, Hahn HJ, Gibert J, Mermillod-Blondin F, Messana G, Notenboom J, Sket B (2008) Incorporation of groundwater ecology in environmental policy. In: Quevauviller P (ed) Groundwater science and policy, chap 11.2. RSC (Royal Soc. of Chemistry), London, pp 671 – 689

Danielson JJ, Gesch DB (2011) Global multi-resolution terrain elevation data 2010 (GMTED2010). US Geol Surv Open-File Rep 2011 – 1073, 26 pp

Datry T, Malard F, Niederreiter R, Gibert J (2003) Video-logging for examining biogenic structures in deep heterogeneous subsurface sediments. CR Acad Sci Paris Life Sci 326:589 – 597

Deharveng L, Stoch F, Gibert J, Bedos A, Galassi D, Zagmajster M, Brancelj A, Camacho A, Fiers F, Martin P, Giani N, Magniez G, Marmonier P (2009) Groundwater biodiversity in Europe. Freshw Biol 54:709 – 726

Dole-Olivier MJ, Castellarini F, Coineau N, Galassi DMP, Martin P, Mori N, Valdecasas A, Gibert J (2009a) Towards an optimal sampling strategy to assess groundwater biodiversity: comparison across six European regions. Freshw Biol 54:777 – 796

Dole-Olivier MJ, Malard F, Martin D, Lefebure T, Gibert J (2009b) Relationships between environmental variables and groundwater biodiversity at the regional scale. Freshw Biol 54:797 – 813

Durr H, Meybeck M, Durr SH (2005) Lithologic composition of the Earth ' s continental surfaces derived from a new digital map emphazing riverine material transfer. Glob Biogeochem Cycles 19:GB4S10. doi:10.1029/2005GB002515

ESRI (2010) ArcGIS desktop: release 9.3. Environmental Systems Research Institute, Redlands, CA FAO, IIASA, ISRIC, ISS-CAS, JRC (2012), Harmonized world soil database (version 1.2), FAO, Rome

Farr TG, Kobrick M (2000) Shuttle radar topography mission produces a wealth of data. EOS Trans AGU 81:583 – 585

Foody GM (2002) Status of land cover classification accuracy assessment. Remote Sens Environ 80:185 – 201

Foody GM (2007) Map comparison in GIS. Progr Phys Geog 31(4):439 – 445

Foster SSD, Chilton PJ (2003) Groundwater: the processes and global significance of aquifer degradation. Phil Trans R Soc Lond B 358:1957 – 1972

Gaffin SR, Rosenzweig C, Xing X, Yetman G (2004) Downscaling and geo-spatial gridding of socio-economic projections from the IPCC Special Report on Emissions Scenarios (SRES). Glob Environ Chang 14:105 – 123

GBIF (2001) Global biodiversity information facility. http://www.gbif.org. Accessed March 2013

Gibert J, Danielopol DL, Stanford JA (1994a) Groundwater ecology. Academic, New York

Gibert J, Stanford JA, Dole-Olivier MJ, Ward JV (1994b) Basic attributes of groundwater ecosystems and prospects for research. In: Gibert J, Danielopol DL, Stanford JA (eds) Groundwater ecology. Academic, New York, pp 7 – 40

Gilbrich WH (2000) International hydrogeological map of Europe. Waterways J 19:11

Gleeson T, Smith L, Moosdorf N, Hartman J, Durr HH, Manning AH, Van Beel LPH, Jellinek AM (2011) Mapping permeability over the surface on Earth. Geophys Res Lett 38:L02401. doi:10.1029/2010GL045565

Gogu RC, Carabin G, Hallet V, Peters V, Dassargues A (2001) GIS-based hydrogeological databases and groundwater modelling. Hydrogeol J 9:555 – 556

Griebler C, Stein H, Kellermann C, Berkhoff S, Brielmann H, Schmidt S, Selesi D, Steube C, Fuchs A, Hahn HJ (2010) Ecological assessment of groundwater ecosystems: vision or illusion? Ecol Eng 36:1174 – 1190

Grunwald S, Thompson JA, Boettinger JL (2011) Digital soil mapping and modelling at continental scales: finding solutions for global issues. Soil Sci Soc Am J 75:1201 –1213

Hahn HJ (2009) A proposal for an extended typology of groundwater habitats. Hydrogeol J 17:77 – 81

Hahn HJ, Fuchs A (2009) Distribution patterns of groundwater communities across aquifer types in south-western Germany. Freshw Biol 54:848 – 860

Hancock PJ, Boulton AJ, Humphreys WF (2005) Aquifers and hyporheic zones: towards an ecological understanding of groundwater. Hydrogeol J 13:98 – 111

Hancock PJ, Hunt RJ, Boulton AJ (2009) Hydrogeoecology the interdisciplinary study of groundwater dependent ecosystems. Hydrogeol J 17:1 – 3

Hiscock KM (2011) Groundwater in the 21st century: meeting the challenges. In: Jones JAA (Ed) Sustaining groundwater resources, International Year of Planet Earth, Springer, Heidelberg, Germany, pp 207 – 225

Hof C, Brändle M, Brandl R (2008) Latitudinal variation of diversity in European freshwater animals is not concordant across habitat types. Glob Ecol Biogeogr 17:539 – 546

Hollis JM, Holman IP, Burton RGO, Willers CM,Pavelley D (2002) A digital data set of European water resources at 1:500.000. Report on behalf of the European Commission and the European Crop Protection Association, 29 pp. http://eusoils.jrc.ec.europa.eu/ESDB_Archive/groundwater/docs/EUGWFR.pdf. Accessed March 2013

Humphreys WF (2009) Hydrogeology and groundwater ecology: does each inform the other? Hydrogeol J 17:5 – 21

IAH (International Association of Hydrogeologists) (1983) International legend for hydrogeological maps. Document SC-84/WS/7. UNESCO, Paris, 49 pp

IGRAC (2005) Global groundwater regions. International Groundwater Resources Assessment Centre, Delft, The Netherlands. Available at http://www.un-igrac.org/publications/329. Accessed March 2013

Illies J (1978) Limnofauna Europaea, 2nd edn. Fischer, Stuttgart, Germany

IUCN (2012) The IUCN red list of threatened species, spatial data V2010.4. http://www.iucnredlist.org/technical-documents/spatial-data. Accessed March 2013

Jones JAA (2011) Groundwater in peril. In: Jones JAA (Ed) Sustaining groundwater resources, International Year of Planet Earth, Springer, Heidelberg, Germany, pp 1 – 19

Larned ST (2012) Phreatic groundwater ecosystems: research frontiers for freshwater ecology. Freshw Biol 57:885 – 906

Laxton JL, Becken K (1996) The design and implementation of a spatial data base for the production of geological maps. Comput Geosci 22(7):723 – 733

Lehner B, Verdin K, Jarvis A (2008) New global hydrography derived from spaceborn elevation data. EOS Trans AGU 89:93 –95

Liu C, Frazier P, Kumar L (2007) Comparative assessment of the measures of thematic classification accuracy. Remote Sens Environ 107:606 – 616

Loveland TR, Reed BC, Brown JF, Ohlen DO, Zhu Z, Yang L, Merchant W (2000) Development of a global land cover characteristics database and IGBP DISCover from 1km AVHRR data. Int J Remote Sens 21(6 – 7):1303 – 1330

MacDonald AM, Robins NS, Ball DF, Dochartaigh BEO (2005) An overview of groundwater in Scotland. Scott J Geol 41(1):3 – 11

MacDonald AM, Bonsor HC, Dochartaigh BEO, Taylor RG (2012) Quantitative maps of groundwater resources in Africa. Environ Res Lett 7:1 – 7

Malard F (2012) Filling the gap: distribution database of European groundwater crustaceans. BioFresh Newsl 4:7. http://www.freshwaterbiodiversity.eu/tl_ fi les/downloads/Downloads/BioFresh_newsletter_issue_004.pdf. Accessed March 2013

Malard F, Hervant F (1999) Oxygen supply and the adaptations of animals in groundwater. Freshw Biol 41:1 – 30

Malard F, Tockner K, Dole-Olivier MJ, Ward JV (2002) A landscape perspective of surface – subsurface hydrological exchanges in river corridors. Freshw Biol 47:621 – 640

Malard F, Boutin C, Camacho A, Ferreira D, Michel G, Sket B, Stoch F (2009) Diversity patterns of stygobiotic crustaceans across multiple spatial scales in Europe. Freshw Biol 54(4):756 – 776

Masuch-Oesterreich D (2000) A methodical approach to GIS-based hydrogeological mapping. Rev Mex Cienc Geol 17(1):24 – 33

Michel G, Malard F, Deharveng L, Lorenzo D, Sket TB, de Broyer C (2009) Reserve selection for conserving groundwater biodiversity. Freshw Biol 54(4):861 – 876

Monfreda C, Ramankutty N, Foley JA (2008) Farming the planet 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. Glob Biogeochem Cycles 22:GB1022. doi:10.1029/2007GB002947

Morris BL, Lawrence ARL, Lawrence, Chilton PJC, Adams B, Calow RC, Klinck BA (2003) Groundwater and its susceptibility to degradation: a global assessment of the problem and options for management. Early warning and assessment report series, RS. 03 – 3. UNEP, Nairobi, Kenya

New M, Hulme M, Jones P (2002) A high-resolution data set of surface climate over global land areas. Clim Res 21:1 – 25

Nikas K, Antonakos A, Kallergis G, Kounis G (2010) International hydrogeological map of Europe: sheet D6 “ Athina ” . Bull Geol Soc Greece 43(4):1821 – 1830

R Development Core Team (2006) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org. Accessed March 2013

Richts A, Struckmeier WF, Zaepke M (2011) WHYMAP and the groundwater resources map of the world 1:25,000,000. In: Jones JAA (ed) Sustaining groundwater resources, International Year of Planet Earth. Springer, Heidelberg, Germany, pp 159 –173

Rouch R (1986) Sur l ’ écologie des eaux souterraines dans le karst [About the ecology of karstified aquifers]. Stygologia 2:352 –398

Rutledge D (2003) Landscape indices as measures of the effects of fragmentation: can pattern re fl ect process? Doc Science Internal Series 98. Department of Conservation, Wellington, New Zealand

Stassberg G, Maidment DR, Jones NL (2007) A geographic data model for representing ground water systems. Groundwater 45(4):515 – 518

Stehman SV (1999) Comparing thematic maps based on map value. Int J Remote Sens 20(12):2347 – 2366 Stehman SV (2000) Practical implications of design based sampling inference for thematic map accuracy assessment. Remote Sens Environ 72:35 – 45

Stehman SV (2009) Sampling designs for accuracy assessment of land cover. Int J Remote Sens 30(20):5243 – 5272

Stein H, Griebler C, Berkhoff S, Matzke D, Fuchs A, Hahn HJ (2012) Stygoregions: a promising approach to a bioregional classification of groundwater systems. Sci Rep 2(673):1 – 9. doi:10.1038/srep00673

Steube C, Richter S, Griebler C (2009) First attempts toward an integrative concept for the ecological assessment of groundwater ecosystems. Hydrogeol J 17:23 – 35

Stoch F, Galassi DMP (2010) Stygobiotic crustacean species richness: a question of numbers, a matter of scale. Hydrobiologia 653:217 – 234

Stoch F, Artheau M, Brancelj A, Galassi DMP, Malard F (2009) Biodiversity indicators in European groundwaters: towards a predictive model of stygobiotic species richness. Freshw Biol 54(4):745 – 755

Struckmeier WF (2008) Contribution of hydrogeological mapping to water monitoring programmes. In: Quevauvlier PP, Borchers U, Thompson C, Simonard T (eds) The water framework directive: ecological and chemical status monitoring. Wiley, New York, pp 215 – 227

Struckmeier WF, Margat J (1995) Hydrogeological maps: a guide and a standard legend. IAH Publ. 17, Heise, Hannover, 177 pp

Tabor K, Williams JW (2010) Globally downscaled climate projections for assessing the conservation impacts of climate change. Ecol Appl 20:554 – 565

Valett HM, Morrice JA, Dahm CN, Campana ME (1996) Parent lithology, surface-groundwater exchange, and nitrate retention in headwater streams. Limnol Oceanogr 41:333 – 345

Venables WN, Ripley BD (2002) Modern applied statistics with S. Springer, New York

Vogt J, Soille P, Colombo R, Paracchini ML, De Jager A (2007) Development of a pan European river and catchment database. In: Digital terrain modelling, lecture notes in geoinformation and cartography. Springer, Heidelberg, Germany, pp 121 – 144

Vorosmarty CJ, Fekete B, Meybeck M, Lammers RB (2000) A stimulated topological network representing the global system of rivers at 30-minute spatial resolution (STN-30). Glob Biogeochem Cycles 14:599 – 621

Ward TJ, Vanderklift MA, Nicholls AO, Kenchington RA (1999) Selecting marine reserves using habitats and species assemblages as surrogates for biological diversity. Ecol Appl 9(2):691 – 698

Wendland F, Blum A, Coetsiers M, Gorova R, Grif fi oen J, Grima J, Hinsby K, Kunkel R, Marandi A, Melo T, Panagopoulos A, Pauwels H, Ruisi M, Traversa P, Vermooten JSA, Walraevens K (2008) European aquifer typology: a practical framework for an overview of major groundwater composition at European scale. Environ Geol 55:77 – 85

Wodja P, Brouyere S, Dassargues A (2010) Geospatial information in hydrogeological studies. In: Anderson MG, McDonnell JJ (eds) Encyclopedia of hydrological sciences, vol 22. Wiley, Chichester, UK

2.2) Article 2 : Geographic variation in range size and beta diversity of groundwater crustaceans: insights from habitats with low thermal seasonality

Maja Zagmajster [1], David Eme [2], Cene Fišer [1], Diana Galassi [3], Pierre Marmonier [2], Fabio Stoch [3], Jean-Francois Cornu [4] and Florian Malard [2]

[1] SubBioLab, Department of Biology, Biotechnical Faculty, University of Ljubljana, Večna pot 111, p.p. 2995, SI-1000 Ljubljana, Slovenia

[2] Université de Lyon, UMR5023 Ecologie des hydrosystèmes Naturels et Anthropisés, Université Lyon1, ENTPE, CNRS, 6 rue Raphaêl Dubois, 69622 Villeurbanne, France

[3] Department of Life, Health & Environmental Sciences, University of L'Aquila, Via Vetoio, Coppito, I-67100 L'Aquila, Italy

[4] UMR BOREA, Département Milieux et Peuplements Aquatiques, MNHN, CNRS 7208, IRD 207, UPMC, Muséum National d'Histoire Naturelle, 43, rue Cuvier ; 75231 Paris, France

Keywords Habitat heterogeneity, historic climate, climate seasonality, Rapoport effect, groundwater, subterranean biodiversity, Crustacea, Europe, species range, beta diversity.

Article accepté dans *Global Ecology and Biogeography* DOI: 10.1111/geb.12200, le 29 Avril 2014.

ABSTRACT

Aim Three broad mechanisms have been proposed to explain geographic variation in species range size: habitat area / heterogeneity, climate seasonality, and long-term climate variability. However, it has proved difficult to disentangle their relative role, particularly as temperature seasonality often co-varies with the amplitude of long–term temperature oscillations. Here, we shed new light into this debate by providing the first continental-scale analysis of range size and beta diversity in groundwater habitats, where taxa are not exposed to latitudinal variation in temperature seasonality.

Location Europe.

Methods We compiled and mapped occurrence data for 1,570 groundwater crustacean species. Generalized regression models were used to test for latitudinal variation in geographic range size and to assess the relative role of the three broad mechanisms in shaping present-day patterns of range size. We partitioned beta diversity into its spatial turnover and nestedness components and analyzed their latitudinal variation across Europe.

Results Median range size increases with latitude above 43 °N and individual species' range size is positively correlated to latitude, even after accounting for phylogenetic effects. Long-term temperature variability accounted for a substantially higher variation in median range size of groundwater

crustaceans across Europe than precipitation seasonality and habitat heterogeneity, including aquifer area, elevation range, climatic rarity and productive energy. Spatial turnover contributes significantly more to beta diversity in southern regions characterized by stable historic climates than it does in northern Europe.

Main conclusions Our findings add support to the historic climate hypothesis which suggests that patterns of increasing range size and decreasing species turnover at higher latitudes in the Palaearctic region are primarily driven by long-term temperature oscillations rather than by climatic seasonality and the availability and heterogeneity of habitats.

Key words Habitat heterogeneity, historic climate, climate seasonality, Rapoport effect, groundwater, subterranean biodiversity, Crustacea, Europe, species range, beta diversity.

INTRODUCTION

Identifying the drivers of geographic variation in range size at continental to global scales is one of the most challenging issues in macroecology (Morueta-Holme *et al*., 2013; Veter *et al*., 2013). Spatial patterns of species richness and beta diversity (i.e. variation in species composition) are inherently linked to the geographic distribution of range size. Therefore, understanding the latter may be key to explaining variation in the richness and composition of communities across space (Weiser *et al.*, 2007). Three major mechanisms have been proposed to explain spatial patterns of range size: habitat area / heterogeneity, climate seasonality, and long-term climate variability (Morueta-Holme *et al*., 2013). Regions with large areas of suitable habitats offer higher potential for range expansion. Small range size has been associated with higher habitat heterogeneity in three different ways. First, steep climatic gradients along mountain slopes may promote local survival by short distance dispersal during periods of climate change (i.e. the climatic buffer hypothesis; Ohlemüller *et al.*, 2008). Second, areas with rare climates may favor adaptation to unusual climatic conditions (i.e. the climatic rarity hypothesis; Ohlemüller *et al.*, 2008). Third, increasing resource availability in high-energy areas may allow niche position specialists to exploit rare resources and niche breadth specialists to forage on their preferred resources (i.e. the specialization hypothesis; Bonn *et al.*, 2004).
Alternatively, the important role of climatic variability in selecting for large range size has gained support from several studies, but the question of which temporal scale of climatic variability matters most is still a matter of debate (Morueta-Holme *et al.*, 2013; Veter et *al.*, 2013). On one hand, Stevens' principle states that local adaptation to more seasonally variable temperatures select for generalists, thereby enabling the species as a whole to expand over wider climatic range (Stevens, 1989). On the other hand, the "historic climate stability" hypothesis encompasses three non-mutually exclusive explanations that emphasize the role of climatic variability at a much longer time scale (Leprieur *et al.*, 2011). These are the disproportionate extinction of small-range taxa in regions severely affected by cold Pleistocene climate (Rohde, 1996), a stronger selection for generalism and vagility imposed by the increasing amplitude of Milankovitch climatic oscillations (Jansson & Dynesius, 2002) and the differential ability of some taxa to colonize vacant habitats following climatic recovery from the Last Glacial Maximum (LGM, i.e. 21,000 years ago; Price *et al.*, 1997). Isolating one or the other scale of variability is difficult because temperature seasonality often co-varies with the amplitude of long–term climatic oscillations (Veter *et al*., 2013).

Groundwater habitats (i.e. all temporally and permanently water-saturated zones in the subsurface) offer useful case studies for exploring key macroecological issues such as the determinants of spatial patterns of species range size. The mean annual groundwater temperature

closely tracks that of air temperature but seasonal variation of temperature is drastically reduced because the annual ground temperature amplitude decreased exponentially with depth below the soil surface (Freeze & Cherry, 1979). Even though there are depth– dependent differences in temperature among groundwater habitats, thermal seasonality is always less than in the surface, even in shallow subterranean habitats (Culver & Pipan, 2011). In groundwater, intra-annual environmental variability is more likely caused by precipitation seasonality which controls variation in subsurface flow conditions and organic matter supply (Culver & Pipan, 2009). Consequently, thermal seasonality cannot drive geographic variation in range size. Any pattern of increasing species range size at higher latitudes (i.e. the Rapoport effect; Stevens, 1989) should reflect the effect of long–term climatic oscillations, precipitation seasonality and/or habitat heterogeneity.

To date, no studies have attempted to evaluate the relative importance of short- and long-term climatic variability and habitat heterogeneity in shaping continental patterns of range size in a thermally stable environment. In groundwater, historic climate oscillations potentially have a strong effect on spatial patterns of range size because most species show small ranges and presumably have poor dispersal ability (Trontelj *et al.*, 2009). If range size is negatively related to extinction rate (Hugueny *et al.*, 2011), then, many groundwater species must have gone extinct in regions of cold Pleistocene climates. This historic effect can persist for longer in groundwater, because it is less likely to be overwritten by subsequent dispersal phases (Foulquier *et al.*, 2008). Two predictions can be made if historic climate changes have been instrumental in shaping European groundwater patterns of range size and beta diversity. First, long-term climatic variability would contribute substantially more than short-term climatic variability and habitat heterogeneity in explaining geographic variation in range size. Second, the proportion of beta diversity explained by species replacement would be higher in southern regions that were less affected by cold Pleistocene climates (Baselga *et al.*, 2012).

The present study provides the first comprehensive analysis of diversity patterns in the European groundwater fauna. First, we compiled and mapped occurrence data for 1,570 species and subspecies of groundwater obligate crustaceans. Second, generalized regression models were used to test for latitudinal variation in geographic range size and to assess the relative role of the three broad mechanisms in shaping present-day patterns of range size. Finally, we partitioned beta diversity into its spatial turnover and nestedness components and analyzed their latitudinal variation across Europe.

MATERIALS AND METHODS

Distributional data set

We assembled in the European groundwater crustacean data set (EGCD) a total of 21,700 database records collectively representing 12 orders, 46 families, 165 genera and 1,570 species and subspecies of Crustacea (Table 1). Half of the occurrence data were extracted from the European PASCALIS database (Deharveng *et al.*, 2009), the hypogean crustacean recording scheme (United Kingdom, Knight 2012), the distributional checklist of the Italian fauna (Ruffo & Stoch, 2006) and the Berlin museum collection. The other half was from authors' institution data sets which were supplemented by an extensive literature search (i.e. 1,380 literature sources). Species occurrence data from the literature were georeferenced to the highest practicable resolution using the spatial coordinates of sampling sites or those of the nearest built-up areas. For each locality, precision of spatial coordinates was indicated as classes (100 m, 1 km and 10 km). In some cases, species distribution maps from the literature were scanned and georectified and the coordinates of occurrence points were computed in ArcGIS 9.3 software (ESRI, 2010). Species names and distributions were checked by taxonomic experts and spurious occurrences were excluded from the data set. For the sake

of exhaustiveness, we included in the data set some undescribed species which were recognized as new to science by experts based on morphological and/or molecular identifications.

Measures of range size and beta diversity

Range sizes of individual species were measured by projecting occurrence data in a Lambert azimuthal equal area projection system (ETRS 1989). Maximum linear extent (MLE) – defined as the straight-line distance between the two most distant known localities – was used as a measure of extent of

Table 1: Orders of crustaceans and their respective numbers of families, genera and species and subspecies included in the distributional data set of groundwater crustaceans in Europe.

Order	Number of families	Number of genera	Number of species and subspecies
Cladocera	1	3	7
Calanoida	1	3	5
Cyclopoida	1	14	172
Gelyelloida	1	1	2
Harpacticoida	7	39	368
Podocopida	7	20	114
Isopoda	6	23	337
Amphipoda	13	33	438
Bathynellacea	2	20	106
Thermosbaenacea	2	2	2
Mysidacea	3	3	3
Decapoda	2	4	16
Total	46	165	1570

occurrence. Area of occupancy (AOO) was quantified by counting the number of 20 × 20 km grid cells in which a species occurred. To analyze the latitudinal patterns of species richness and median range size of all species, we used the grid of 0.9° latitude cells provided by the EDIT geoplatform (Sastre *et al.*, 2009). Cell area was kept constant (10,000 km^2) all over the grid by adjusting the longitudinal divisions between adjacent cells in each latitudinal band. Coastal cells with < 20% of land area were excluded from the grid (except for islands). The grid had 701 cells, 494 of which contained at least one species occurrence. We computed the median range size of species contained in each cell using both measures of range size (MLE and AOO). Richness was defined as the number of species and subspecies contained in each cell because subspecies were considered as distinct phylogenetic units (see Agapow *et al.*, 2004).

Pairwise measures of beta diversity as implemented by Baselga (2012) were used to assess the increase in crustacean assemblage dissimilarity (i.e. high beta diversity) with geographic distance. Total beta diversity as represented by the Jaccard index of dissimilarity ($\beta_{jac,}$ ranging from 0 [perfect similarity] to 1 [perfect dissimilarity]) was partitioned into its turnover ($\beta_{jtu,}$ differences in composition caused by species replacements) and nestedness-resultant ($\beta_{jne,}$ differences in species composition caused by species losses or gains) components.

Multiple-site measures of beta diversity were used to examine the spatial pattern of variation in the intra-regional turnover (β_{JTU}) and nestedness-resultant (β_{JNE}) components of beta diversity in Europe (Baselga, 2012). Following Svenning *et al.* (2011), larger regions for this analysis

corresponded to squares provided by the EDIT Geoplatform (Sastre *et al.*, 2009), all having the same area (250,000 km^2) and latitudinal range (4.5°).

Predictors of median range size

We used seven predictors to test for the relative importance of the three mechanisms in shaping spatial patterns of range size. Habitat heterogeneity was represented by aquifer area, elevation range, climatic rarity and productive energy. Aquifer area was obtained for each grid cell by calculating the total area of aquifer available within a 1000-km radius around the focal cell. The spatial distribution of aquifers was derived from the European map of groundwater habitats (Cornu *et al.*, 2013). Elevation range was calculated as the highest difference in elevation between any two locations in a cell using elevation data from the pan-European River and catchment database (Vogt *et al.*, 2007). For climatic rarity, we followed Morueta-Holme *et al.* (2013) and extracted and averaged for each grid cell data on mean annual temperature and all 8 bioclimatic precipitation layers of the WorldClim data set, except precipitation seasonality (30 arc-second resolution; Hijmans et al. 2005). Then, a normalized principal component analysis (PCA) was performed on grid-cell averaged data, log-transformed when appropriate to approximate a normal distribution. The first two PCA axes accounted for 86.8 % of the climatic variance. They were used to calculate the average Euclidian distance in climatic space between each cell and all neighboring cells within a radius of 500 km. High average distance values correspond to cells showing rare climates relative to their neighboring cells (Morueta-Holme *et al.*, 2013). To test for the effect of productive energy on the distribution of range size, we extracted data on actual evapotranspiration (AET) from the 30 arc-second CGIAR data set (Trabucco & Zomer, 2010). To measure long-term climatic variability, we used temperature and precipitation anomalies, defined as the differences in mean annual temperature and annual precipitation between the present and LGM (Leprieur *et al.*, 2011). Data for LGM were extracted from two global circulation models, CCSM and MIROC2 (Hijmans *et al.*, 2005) and values from both were averaged to account for variation between models. To assess the role of short-term climate variability, we used precipitation seasonality as a surrogate of present intra-annual environmental variability.

Statistical analyses

Latitudinal pattern of range size

Relationships between cell average of median range size per 0.9° latitudinal band and latitude were assessed by means of ordinary least squares models (OLS) and generalized additive models (GAM) to account for curvilinear relationships. The best models were selected as those having the minimum Akaike information criterion (AIC). Generalized least squares (GLS) were used to test for differences in median range size, while accounting for unequal variances among latitudinal bands. All statistical analyses were performed with R software (R Development Core Team, 2011) using mgcv package for GAM (Wood, 2011) and nlme package for GLS (Pinheiro *et al.*, 2011).

To assess the effect of sampling bias on the observed patterns of median range size and species richness, we used the number of database records per cell as a surrogate for sampling effort (Ballesteros-Mejia *et al*., 2013). A record was defined as a unique combination of species, locality and date. Three subsets of cells containing at least 5, 10 and 20 records were used to compute for each cell an index of sampling completeness (SC), defined as the ratio of observed to *Chao1* estimated species richness (Soberón e*t al.*, 2007, but see Zagmajster *et al.*, 2010). We retained only those cells with SC > 0.8 to check for the robustness of the relationship between cell average of median range size or species

richness per latitudinal band and latitude. *Chao1* richness estimates were computed using the vegan package (Oksanen *et al.*, 2012).

Relationships between measures of range size and latitude were also assessed using species as individual points because latitudinal band methods suffer from pseudoreplication (Ruggiero & Werenkraut, 2007). The relationship between range size and the latitudinal midpoint of species' ranges was tested with OLS and GLS models using different structures of residual variance. Midpoint of species' ranges was included in the models in its linear and quadratic forms. We also performed phylogenetic generalized least squares (PGLS) to account for phylogenetic non-independence among species. A nested grouping structure based on taxonomy (order, family, genus, and species) was used because a comprehensive molecular phylogeny of groundwater crustaceans was not available. Taxonomic distance was assumed to be equal among hierarchical levels.

We tested whether the evolution of range size and latitudinal midpoint was best fitted by: 1) a Brownian motion model; 2) a Brownian motion model with the optimized lambda parameter, and; 3) the Ornstein-Uhlenbeck (OU) model. The best evolution model was selected according to minimum AIC and the amount of explained variation was estimated according to Nagelkerke (1991). Taxonomic distances were computed with the vegan package and selection of evolution model and PGLS were performed with ape (Paradis *et al.*, 2004) and nlme packages.

Drivers of spatial variation in median range size

To test the role of the three broad mechanisms on the spatial pattern of range size, we performed multiple OLS models using the 7 predictors. Then, variance partitioning (Legendre & Legendre, 1998) was used to estimate the unique and shared contributions of the three mechanisms. Models were performed using the full data set (n = 494 cells) and the three data subsets comprising only those cells with SC > 0.8. We also analyzed the data using simultaneous autoregressive (SAR) models because spatial autocorrelation in the residuals of OLS models could potentially affect the estimation of parameters and their statistical significance (Dormann *et al.*, 2007). The statistical procedure used to perform OLS and SAR models and variance partitioning is described in Appendix S4.

Latitudinal pattern of beta diversity

Multiple-cell dissimilarity measures were used to assess variation in the turnover (β_{JTU}) and nestedness components (β_{JNE}) of intra-regional beta diversity across 250,000 km^2 squares in Europe. Differences in the number of cells among squares (mean: 18 ± 5 cells; range: 11-25 cells) were controlled for by re-sampling 11 cells from each square 1000 times. Then, a GLS model with logit transformation was used to test for latitudinal variation in the turnover component of regional beta diversity.

Pairwise measures of turnover (β_{jtu}) and nestedness-resultant dissimilarity (β_{jne}) were used to assess the increase in crustacean dissimilarity with geographic distance within three bands of equal latitudinal range (35-41.3°N, 41.3-47.6°N and 47.6-53.9°N) (Svenning *et al.*, 2011). Boundaries of the middle band were delineated as to include the most species-rich cells and the southern and northern bands were defined so that they had the same latitudinal range than the middle band (i.e. 6.3°). In each band, significance of Pearson correlation coefficient for relationship between dissimilarity and distance was computed by means of Mantel permutation tests (1000 permutations). The frequency distributions of slopes and intercepts was estimated for each band by bootstrapping (n = 1000). To test

for the significance of one parameter being larger in one band than in the other, we computed the probability of obtaining the opposite results by chance by comparing the estimated distributions of parameters. Probabilities were adjusted with Holm corrections for multiple comparisons among bands.

We also performed Mantel correlograms using Pearson's correlation coefficient between β_{jtu} and geographic distance in order to determine the geographic distance at which the Mantel correlation coefficient was no longer significant or became negative. A total of seven distance classes were used to allow a similar and sufficient number of pairwise comparisons between cells to be considered in each correlation coefficient calculation. Multiple and pairwise beta diversity indexes were computed with the R betapart package (Baselga & Orme, 2012) and Mantel tests, correlograms and bootstrapping procedure were performed using the vegan and boot (Canty & Ripley, 2011) packages.

RESULTS

Latitudinal pattern of range size

Hereafter, the term range size refers to as MLE and AOO: statistical results are given in the main text for both measures but all graphical displays for AOO are provided in Supporting Information (Appendices S1, S2). Cell average of median range size per latitudinal band was positively related to latitude but GAMs notably improved model fit over an OLS model (Figs.1 & 2, Appendices S1, S2). There was a threshold at approximately 43 degrees of latitude above which averaged range size markedly increased with latitude. Median range size did not differ among bands below the 43rd parallel (Fig. 2, Appendix S1). The pattern of increasing median range size at higher latitudes was confirmed when restricting the analysis to the cells with high sampling completeness (Appendix S3). Species richness exhibited a unimodal pattern with latitude and the ridge of high species richness at latitudes ranging from ca 42° to 46° N was still apparent when using estimated species richness of cells with SC >0.8 (Fig. 1, Appendices S1, S3).

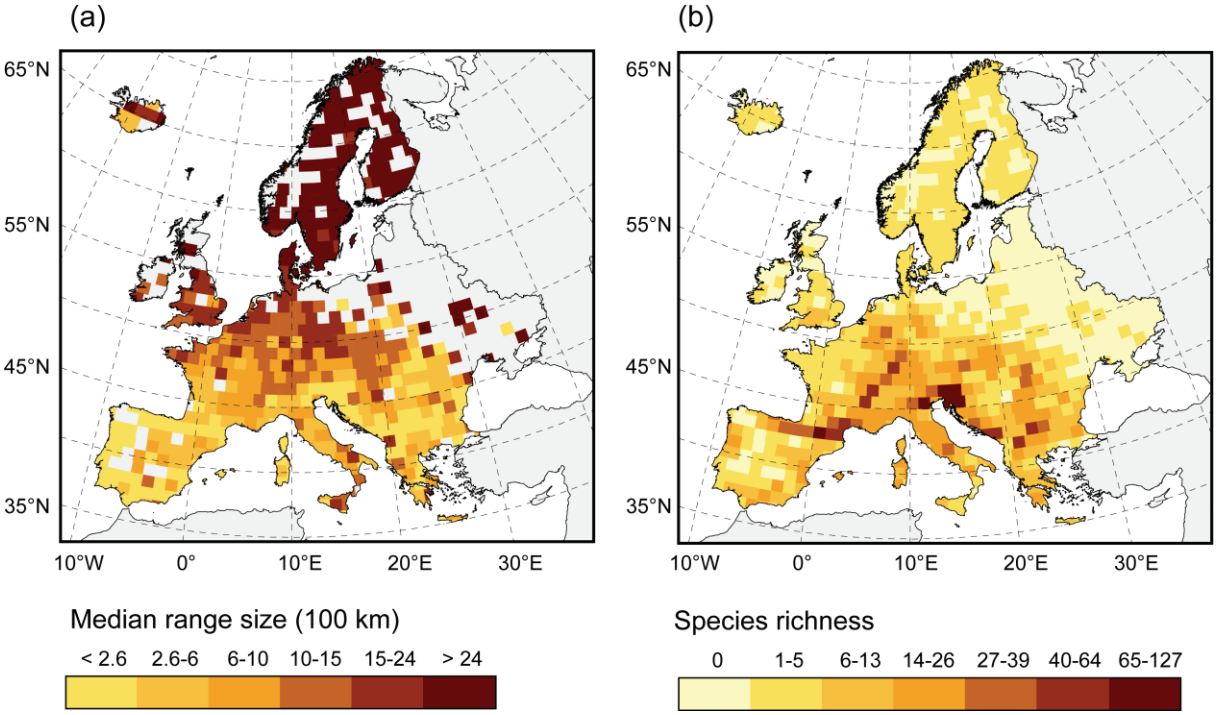


Figure 1: Geographic distribution of median range size (maximum linear extent) of all species (a) and species richness (b) for the European groundwater crustacean fauna. Class delimitation is based on Jenks natural breaks of the underlying distributions. Cell area is 10,000 km^2.

Ordinary least squares models incorporating a quadratic term showed that species range size increased significantly with latitude (Fig. 2, Table 2, Appendix S2). However, GLS models allowing variation in the residuals to increase exponentially as a function of latitude notably improved model fitting to the data (Table 2), indicating that small- and large-range species occurred together at higher latitudes. Yet, the most northern regions (latitude > 54°N) were exclusively colonized by a few widely-distributed species. The relationship between range size and latitude was still significant when the phylogenetic non-independence of species was taken into account using PGLS (Table 2). Incorporating an exponential variance structure significantly improved PGLS fit, proving evidence that range size variation among species increased exponentially with latitude.

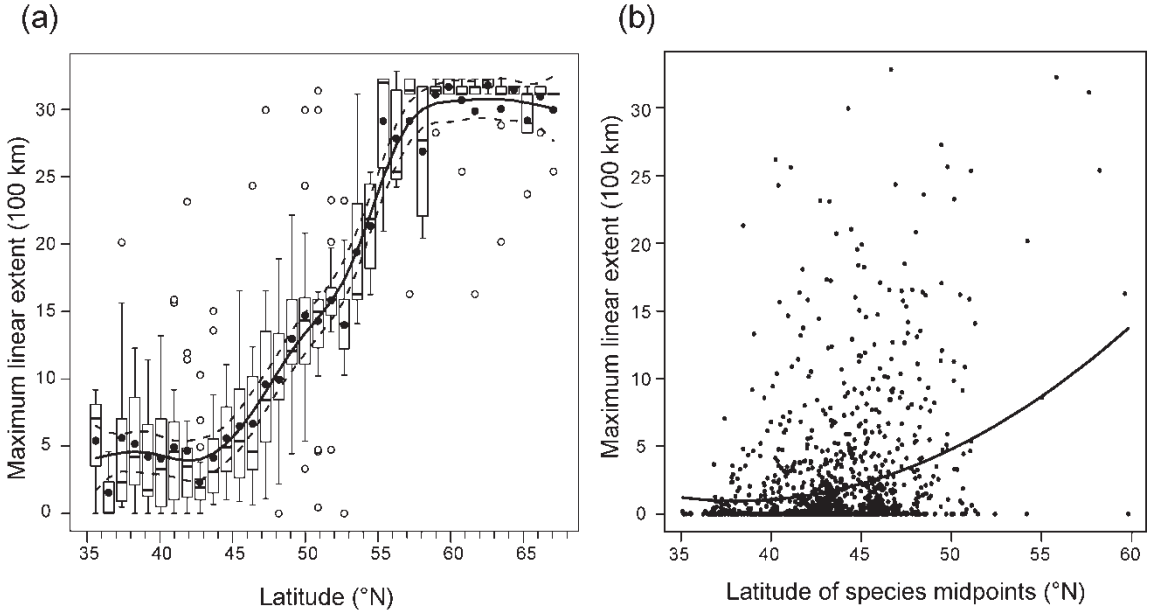


Figure 2: (a) Relationship between median range size (maximum linear extent) per latitudinal band and latitude. Black horizontal bars, black dots, and boxes show the median, average, and interquartile range, respectively, for 0.9° latitudinal bands. The maximum length of each whisker is up to 1.5 times the interquartile range and open circles represent outliers. Continuous black lines represent the fit of a generalized additive model to the averages of latitudinal bands and dashed lines show 95% confidence intervals (see Appendix S1). (b) Relationship between maximum linear range extent of species (n = 1568 species; 2 species endemic to Iceland were excluded) and latitudinal midpoint of species. The black line represents the fit of phylogenetic generalized least squares using the best evolution model (Ornstein-Uhlenbeck), a quadratic term for latitudinal midpoint and an exponential variance structure for the residuals (see Table 2).

Drivers of spatial variation in median range size

The best OLS and SAR models for the full data set explained 71.8 and 81.3% of the overall variance in median range size per cell, respectively (Table 3). Among the three hypotheses, long term climatic variability had by far the highest unique contribution (20.8 to 36.1%) to spatial variation in range size, regardless of the models and data sets used. The historical climate variability hypothesis also shared a substantial amount of variance with habitat heterogeneity, which by itself explained only a small fraction of variation in range size. Median MLE per cell was positively related to temperature anomaly (R^2 = 0.64, n = 494 cells, p < 0.0001). The latter predictor had an AICc weight of 1 in all models and data sets, whereas the weights of precipitation anomaly and seasonality were in most cases lower than 0.5 (Appendix S4). This indicates that long-term temperature variability was more

important than long- and short-term precipitation variability in shaping the distribution of median range size.

Latitudinal pattern of beta diversity

Intra-regional beta diversity as measured by the Jaccard index of multiple-cell dissimilarity was >0.90 in all squares (mean 0.95 ±0.019, n = 19) and the contribution of turnover to beta diversity (β_{JTU} = 94.6%±4) was consistently larger than that of the nestedness-resultant dissimilarity (β_{JNE} = 5.4% ±4) (Fig. 3a). However, the GLS model showed that the proportion of beta diversity explained by species replacement was significantly higher in squares of the two most southern bands than in those of the most northern band ($p < 0.02$, adjusted with Holm corrections, Fig. 3b).

Table 2: Results of ordinary least square, generalized least square and phylogenetic generalized least square models between species' maximal linear range extent and latitudinal midpoint of species' ranges (i.e. mean latitude; n = 1568 species, 2 species endemic to Iceland were excluded). $p < 0.01$ for all parameter estimates.

Model	Parameter	Parameter estimate	Standard error	t	R^2	AIC[e]
OLS [a]	Intercept	7592.9	1084.4	7.0	0.11	23 224
	Linear	-373.6	49.3	-7.6		
	Quadratic	4.7	0.6	8.3		
GLS [b]	Intercept	4249.5	1215.4	3.5	0.23	23 008
	Linear	-216.7	57.8	-3.8		
	Quadratic	2.8	0.7	4.1		
PGLS [c]	Intercept	7380.7	1099.2	6.7	0.10	23 175
	Linear	-362.9	49.7	-7.3		
	Quadratic	4.5	0.6	8.1		
PGLS [d]	Intercept	4040.8	1218.2	3.3	0.23	22 941
	Linear	-206.9	58.0	-3.6		
	Quadratic	2.7	0.7	3.9		

[a] Ordinary least squares
[b] Generalized least squares with an exponential variance structure for the residuals
[c] Phylogenetic generalized least squares with the best evolution model (Ornstein-Uhlenbeck)
[d] Phylogenetic generalized least squares with an exponential variance structure for the residuals and the best evolution model (Ornstein-Uhlenbeck)
[e] Akaike information criterion

Pairwise measure of turnover (β_{jtu}) increased with geographic distance in all latitudinal bands (Fig. 4a-c), whereas the slope was negative for the nestedness-resultant dissimilarity (Fig. 4d-f). Bootstrapped slopes were significantly different across all pairs of bands ($p < 0.001$). Bootstrapped intercepts for β_{jtu} were significantly higher in the two most southern bands ($p < 0.001$) than in the most northern band and the lowest slopes occurred for the two most southern bands (Fig. 4a-c). This implied that turnover was higher in east-west direction in southern Europe. The most northern band had by far the highest intercept ($p < 0.001$) and the steepest slope for β_{jne} (Fig. 4d-f).The Mantel correlation for β_{jtu} was no longer significant or became negative when distance between cells exceeded 500 km (Fig. 4g-i). This implied that the increase in turnover with geographic distance was typically not linear because species replacement levels off for distance > 500 km.

DISCUSSION

Striking features of groundwater crustacean diversity patterns

The present study provides one of the few continental–scale analyses of diversity patterns in a seasonally stable habitat (Rex *et al.*, 1993). The European groundwater crustacean data set (EGCD) is the most comprehensive and updated compilation of data on groundwater species occurrences at a continental scale. It contains two times more species of groundwater crustaceans than the *Limnofauna Europaea*, which was recently used to compare European patterns of species richness and beta diversity among taxa adapted to lotic, lentic and groundwater habitats (Hof *et al.*, 2008). Moreover, we used species occurrence data from georeferenced localities rather than approximate range maps or species lists for broad geographic regions because groundwater species tend to exhibit a high degree of range porosity. Characterizing geographic patterns of diversity by overlaying range maps would have led to overestimate species richness and underestimate beta diversity, more particularly in regions of high habitat heterogeneity and fragmentation (Hurlbert & White, 2005). Rather, patterns were obtained directly by intersecting occurrence data with a grid of cells having the same area and latitudinal range, while only one of these two parameters has usually been kept constant in most macroecological studies (Sastre *et al.*, 2009).

Table 3: Unique and shared contributions of habitat heterogeneity (H), long term climatic variability (L) and short tern climatic variability (S) to spatial variation in median range size (maximum linear extent). In abbreviations, colons denote shared variance between hypotheses. Ordinary least squares (OLS) and simultaneous autoregressive (SAR) models were performed on the full data set (n = 494 cells) and the three data subsets comprising only those cells with sampling completeness >0.8 (see Appendix S4 for details of model results). For SAR models, the explained variance including the spatial component is indicated in parentheses.

Model	Data set	Explained variance (%)	Variance partitioning (%)						
			Habitat heterogeneity (H)	Long term climatic variability (L)	Short term climatic variability (S)	H:L	L:S	H:S	H:L:S
OLS	494 cells	71.8	5.7	28.8	0.2	36.7	1.2	2.0	-2.8
	147 cells	54.8	3.8	25.2	-0.3	23.2	4.1	0.9	-2.1
	115 cells	48.8	1.4	20.8	-0.4	18.1	6.8	0.0	2.1
	83 cells	54.7	6.8	25.7	0.3	18.9	3.1	-1.0	0.8
SAR	494 cells	67.6 (81.3)	4.2	36.1	0.5	26.5	-10.6	0.1	10.9
	147 cells	55.4 (72.6)	4.3	23.7	-0.4	24.1	5.1	1.1	-2.6
	115 cells	51.1 (70.1)	2.4	23.2	-0.2	16.2	7.4	0.2	1.9
	83 cells	54.9 (73.4)	6.8	28.8	-2.5	14.8	7.5	0.8	-1.6

Several striking features of groundwater biodiversity patterns emerge from the analysis of the EGCD. First, there is a consistent positive latitudinal trend in range size, regardless of the methods used. Second, beta diversity of groundwater crustaceans in Europe is mainly caused by spatial turnover but nestedness–resultant dissimilarity contributes significantly more to beta diversity in northern Europe than it does in southern Europe. Third, species–rich cells occur along a relatively

narrow latitudinal band (ca 42 to 46° N) extending in east–west direction from Spain to Bulgaria, which is thereafter referred to as the ridge of high species richness (Culver *et al.*, 2006). Even though sampling bias in large scale databases is unavoidable, the ridge of high species richness and the latitudinal patterns of increasing range size are still apparent, while accounting for variable sampling effort across cells. Yet, we acknowledge that species inventory in groundwater is far from being complete as indicated by the proportion of cells with high sampling completeness.

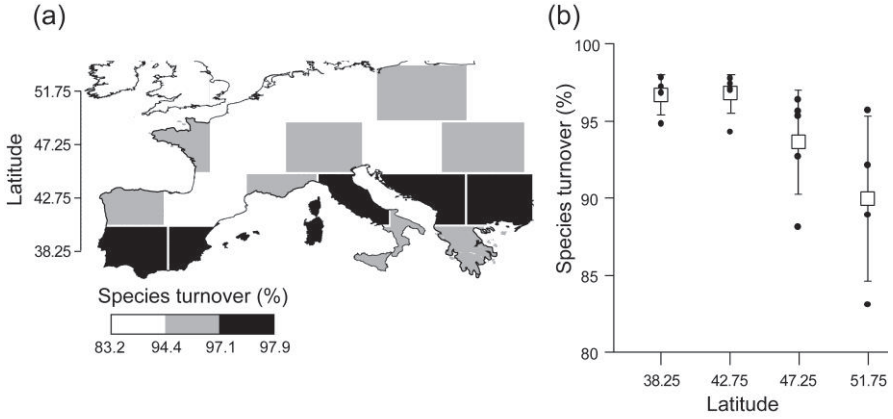


Figure 3: (a) Spatial variation of regional groundwater crustacean species turnover in Europe. Multiple-site species turnover within regions (250,000 km^2 squares) was calculated by re-sampling 11 cells of each square 1000 times. (b) Plot of mean regional species turnover versus latitude. White squares and whiskers show means and standard deviations, respectively.

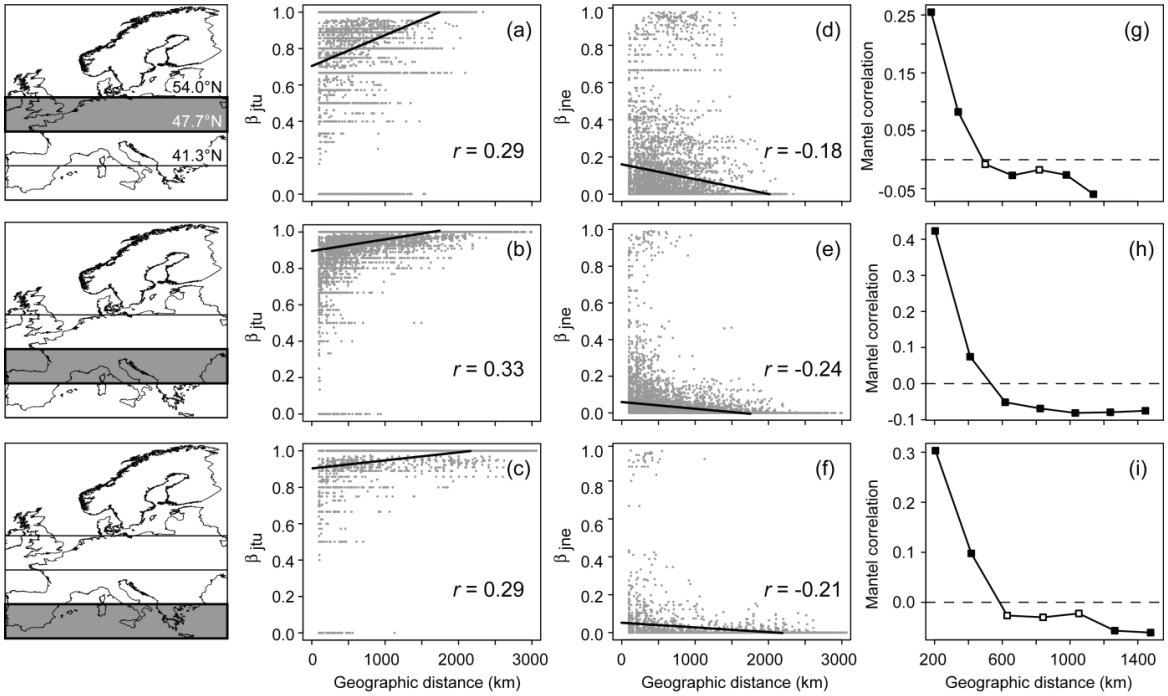


Figure 4: Relationships (linear fits) between groundwater crustacean species turnover (β_{jtu}: a-c), nestedness-resultant dissimilarity (β_{nes}: d-f) and geographic distance, and Mantel correlograms of the Pearson's correlation coefficient between β_{jtu} and distance (g-i) within three latitudinal bands in Europe. All Pearson correlation coefficients (r) in panels a-f were significant (Mantel tests; $p < 0.01$). Filled and open squares in panels g-i indicate significant and non significant Mantel correlation coefficients, respectively.

Patterns and drivers of range size

Our first prediction that geographic variation in range size of groundwater crustaceans would be primarily driven by long-term climatic variability was supported by the results of generalized regression models and variance partitioning. Our findings add support to the historic climate hypothesis which suggests that patterns of range size in the Palaearctic and Nearctic regions might primarily be driven by climatic variability acting at much longer time scales than the seasonal scale (Rohde, 1996; Jansson & Dynesius, 2002). Contrary to Stevens' predictions (Stevens, 1989), we show that the groundwater crustacean fauna exhibits a Rapoport effect, while it is not exposed to increasing temperature seasonality at higher latitudes. Cell average of median range size per latitudinal band increased abruptly above a latitudinal threshold of 43 °N in regions that experienced large amplitude of Quaternary climate oscillations (temperature anomaly between present and LGM > 5°C). Notably, this latitudinal threshold roughly coincides with the «Holdhaus line», which marks the northern distribution of many cave and edaphic blind beetle species (Drees *et al.*, 2010). The positive latitudinal increase in range size for the band method ($r = 0.95$) and the independent point method ($r = 0.48$) was higher than the average increase reported by Ruggiero and Werenkraut (2007) in their meta-analysis of the Rapoport effect ($r = 0.30$ from a total of 49 studies). The individual point method provided a weaker positive latitudinal trend in range size than the band method because variation in range size among species increased with latitude, thereby leading to a greater dispersion in the data.

Long-term climatic variability accounted for a substantially higher variation in median range size of groundwater crustaceans across Europe than short-term climatic variability and habitat heterogeneity. The overriding influence of historic climates was essentially attributed to long-term variation in temperature rather than aridity because the explanatory power of temperature anomaly (AICc weight) was consistently higher than that of precipitation anomaly. The comparatively low explanatory power of aquifer area, elevation range, climatic rarity and productive energy might indicate that the small range size of most groundwater crustaceans is primarily determined by their poor intrinsic dispersal capacities rather than by the availability and heterogeneity of habitats. Recent studies also showed that geographical variation in species range size was primarily determined by long- or short-term climate stability in the Nearctic (Veter *et al.*, 2013), whereas habitat heterogeneity gained in importance in the Neotropical region (Morueta-Holme *et al.*, 2013). In addition, our findings suggest that patterns of range size in the Palaearctic region might primarily be driven by climatic variability acting at much longer time scales than the seasonal scale (Rohde, 1996; Jansson & Dynesius, 2002).

We suggest that the latitudinal increase in groundwater species range size reflects the influence of long-term climatic variability on the trade-off between dispersal ability and ecological specialization (Jansson & Dynesius, 2002). In the absence of thermal seasonality, larger temperature oscillations at higher latitudes might prevent strong thermal niche narrowing. Populations maintaining larger thermal niche breadths can more easily colonize vacant habitats at northern latitudes via shallow subsurface dispersal pathways. Dispersal might further prevent local adaptation due to maladaptive gene flow among populations (Räsänen & Hendry, 2008). Although this scenario needs testing, it is consistent with findings from a recent study showing that individual populations of groundwater species that have expanded over a wide climatic range in Europe could each maintain a wide thermal tolerance breadth although they experienced little seasonal variation of temperature in their natural habitats (Mermillod-Blondin *et al.*, 2013).

Patterns of beta diversity

The partitioning of beta diversity into its turnover and nestedness-resultant components supported our second prediction by revealing a pattern of increasing species replacement with decreasing latitude. Although different processes can lead to species replacement, this finding is consistent with a historic legacy of Pleistocene climate in the composition of the European groundwater crustacean fauna. Indeed, the contribution of species turnover to beta diversity is predicted to be higher in regions characterized by stable climates where speciation events can accumulate over time (Baselga *et al.*, 2012; Dobrovolski *et al.*, 2012). In contrast, higher nestedness-resultant dissimilarity in northern regions experiencing large climatic oscillations can reflect groundwater species losses due to increased extinction rates and colonization during interglacials. Comparison of beta diversity–geographic distance relationships across latitudinal bands also showed that species replacement was higher in the two most southern bands whereas nestedness-resultant dissimilarity was highest in the most northern band. However, postglacial colonization cannot be from distant Mediterranean and continental refuges because constrained dispersal from Mediterranean and eastern faunal source areas to northern sink areas should have resulted in a pattern of increasing nestedness-resultant dissimilarity with distance (Svenning *et al.*, 2011). Instead, we found that nestedness-resultant dissimilarity decreased with geographic distance in all bands. Moreover, a strong nestedness-resultant dissimilarity pattern can only occur if assemblages belong to the same species pool. Yet, Mantel correlograms and the relationships between species turnover and geographic distance indicated that groundwater crustacean assemblages were almost entirely replaced within distance < 500 km. This finding corroborates studies showing that ecoregions are considerably smaller ($<10^5$ km^2) in groundwater than in surface water (Stein *et al.*, 2012). Our results are consistent with a recent phylogeographic study that documented the important role of dispersal events from multiple groundwater refugia located at the northern periphery of the Alps during postglacial colonization (Eme *et al.*, 2013).

Documenting and understanding continental-scale patterns of diversity in environments lacking strong seasonality allow disentangling the role of short- and long-term climatic variability in shaping the geographic distribution of range size. Beyond Stevens' hypothesis, climate seasonality is a central element both in Janzen's model of diversification in tropical taxa (Janzen, 1967; but see also Quintero & Wiens, 2013) as well as in the more recent climate-mediated dispersal–ecological specialization trade-off proposed by Jocque *et al.* (2010). In contrast, the ORD hypothesis (orbitally forced range dynamics; Jansson & Dynesius, 2002) emphasized the significance of selection pressures on specialization and vagility induced by climate variability at time scales well beyond seasonal changes. Groundwater organisms can also offer a useful case study for testing whether long-term climatic oscillations can mediate the dispersal–ecological specialization trade-off, which is ultimately thought to drive large-scale diversity patterns.

ACKNOWLEDGEMENTS

This research was funded by the European programme BioFresh (No. 226874), the French programme DEEP (No. ANR 08JCJC012001) and the PHC Proteus programme (No. 31199UM). We thank A.I. Camacho for checking distributional data for Bathynellacea, J. Notenboom for providing coordinates of localities in Spain, N. Bailly for developing the Stygofauna Mundi database, and C.O. Coleman for providing data from Berlin museum collection. CF and MZ were supported by Slovenian Research Agency (program P1-0184). We thank the editors and two anonymous reviewers for their valuable comments on an earlier draft of this manuscript.

REFERENCES

Agapow, P.-M., Bininda-Emonds, O.R.P., Crandall, K.A., Gittleman J.L., Mace G.M., Marshall J.C. & Purvis A. (2004) The impact of species concept on biodiversity studies. *The Quaterly Review of Biology*, **79**, 161–179.

Ballesteros-Mejia, L., Kitching, I.J., Jetz, W., Nagel, P. & Beck, J. (2013) Mapping the biodiversity of tropical insects: species richness and inventory completeness of African sphingid moths. *Global Ecology and Biogeography*, **22**, 586–595.

Baselga, A. (2012) The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. *Global Ecology and Biogeography*, **21**, 1223–1232.

Baselga, A. & Orme, C.D.L. (2012) betapart: an R package for the study of beta diversity. *Methods in Ecology and Evolution*, **3**, 808–812.

Baselga, A., Gomez-Rodriguez, C. & Lobo, J.M. (2012) Historical legacies in world amphibian diversity revealed by the turnover and nestedness components of beta diversity. *PLoS ONE*, **7**, 2, e32341.

Bonn, A., Storch, D. & Gaston, K. J. (2004) Structure of the species–energy relationship. *Proceedings of the Royal Society of London B*, **271**, 1685–1691.

Canty, A. & Ripley, B. (2011) boot: Bootstrap R (S-Plus) functions. R package version 1.3-3. Available at: http://cran.r-project.org/web/packages/boot/index.html.

Cornu, J.-F., Eme, D. & Malard, F. (2013) The distribution of groundwater habitats in Europe. *Hydrogeology Journal*, **21**, 949–960.

Culver, D.C. & Pipan, T. (2009) Biology of caves and other subterranean habitats. Oxford University Press, Oxford, U.K.

Culver, D.C. & Pipan, T. (2011) Redefining the extent of the aquatic subterranean biotiope – shallow subterranean habitats. *Ecohydrology*, **4**, 721–730.

Culver, D.C., Deharveng, L., Bedos, A., Lewis, J.J., Madden, M., Reddell, J.R., Sket, B., Trontelj, P. & White, D. (2006) The mid-latitude biodiversity ridge in terrestrial cave fauna. *Ecography*, **29**, 120–128.

Deharveng, L., Stoch, F., Gibert, J., Bedos, A., Galassi, D., Zagmajster, M., Brancelj, A., Camacho, A., Fiers, F., Martin, P., Giani, N., Magniez, G. & Marmonier, P. (2009) Groundwater biodiversity in Europe. *Freshwater Biology*, **54**, 709–726.

Dobrovolski, R., Melo, A.S., Cassemiro, F.A.S. & Diniz-Filho, J.A.F. (2012) Climatic history and dispersal ability explain the relative importance of turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, **21**, 191–197.

Dorman, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.

Drees, C., Matern, A., von Oheimb, G., Reimann, T. & Assmann, T. (2010) Multiple glacial refuges of unwinged ground beetles in Europe: molecular data support classical phylogeographic models. *Relict Species: Phylogeography and Conservation Biology* (ed. by J.C. Habel and T. Assmann), pp. 199–215.

Eme, D., Malard, F., Konecny-Dupré, L., Lefébure, T. & Douady, C.J. (2013). Bayesian phylogeographic inferences reveal contrasting colonization dynamics among European groundwater isopods. *Molecular Ecology*, **22**, 5685–5699.

ESRI (2010) ArcGIS Desktop: Release 9.3. Redlands, CA: Environmental Systems Research Institute.

Foulquier, A., Malard, F., Lefebure, T., Gibert, J. & Douady, C.J. (2008) The imprint of Quaternary glaciers o the present-day distribution of the obligate groundwater amphipod *Niphargus virei* (Niphargidae). *Journal of Biogeography*, **35**, 552–564.

Freeze, R.A. & Cherry, J.A. (1979) *Groundwater*. Prentice-Hall, Englewood Cliffs.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hof, C., Brändle, M. & Brandl, R. (2008) Latitudinal variation of diversity in European freshwater animals is not concordant across habitat types. *Global Ecology and Biogeography*, **17**, 539–546.

Hugueny, B., Movellan, A. & Belliard, J. (2011) Habitat fragmentation and extinction rates within freshwater fish communities: a faunal relaxation approach. *Global Ecology and Biogeography*, **20**, 449–463.

Hurlbert, A.H. & White, E.P. (2005) Disparity between range map- and survey-based analyses of species richness: patterns, processes and implications. *Ecology Letters*, **8**, 319–327.

Jansson, R. & Dynesius, M. (2002) The fate of clades in a world of recurrent climatic change: Milankovitch oscillations and evolution. *Annual Review of Ecology and Systematics*, **33**, 741–777.

Janzen, D.H. (1967) Why mountain passes are higher in the tropics. *The American Naturalist*, **101**, 233–249.

Jocque, M., Field, R., Brendonck, L. & De Meester, L. (2010) Climatic control of dispersal–ecological specialization tradeoffs: a metacommunity process at the heart of the latitudinal diversity gradient? *Global Ecology and Biogeography*, **19**, 244–252.

Knight, L. (2012) Hypogean Crustacea Recording Scheme. Data available at: http://hcrs.freshwaterlife.org

Legendre, P., Legendre, L. (1998) Numerical Ecology, 2nd ed. Elsevier Science, Amsterdam.

Leprieur, F., Tedesco, P.A., Hugueny, B., Beauchard, O., Dürr, H.H., Brosse, S. & Oberdorff, T. (2011) Partitioning global patterns of freshwater fish beta diversity reveals contrasting signatures of past climate changes. *Ecology Letters*, **14**, 325–334.

Mermillod-Blondin, F., Lefour, C., Lalouette, L., Renault, D., Malard, F., Simon, L. & Douady, C. J. (2013). Thermal tolerance breadths among groundwater crustaceans living in a thermally constant environment. *Journal of Experimental Biology*, **216**, 1683–1694.

Morueta-Holme, N., Enquist, B.J., McGill, B.J., Boyle, B., Jørgensen, Jeffrey, E.O., Peet, R.K., Šímova, I., Sloat, L.L., Thiers, B., Violle, C., Wiser, S.K., Dolins, S., Donoghue II, J.C., Kraft, N.J.B., Regetz, J., Schildhauer, M., Spencer, N. & Svenning, J.-C. (2013) Habitat area and climate stability determine geographical variation in plant species range sizes. *Ecology Letters*, **16**, 1446–1454.

Nagelkerke, N.J.D. (1991) A note on a general definition of the coefficient of determination. *Biometrika*, **78**, 691–692.

Ohlemüller, R., Anderson, B.J., Araújo, M.B., Butchart, S.H.M., Kudrna, O., Ridgely, R.S. & Thomas, C.D. (2008) The coincidence of climatic and species rarity: high risk to small-range species from climate change. *Biology Letters*, **4**, 568-572.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin , P.R., O'Hara R.B., Simpson, G.L., Solymos, P., Stevens, M., Henry, H. & Wagner, H. (2012) Vegan: Community Ecology Package. R package version 2.0-5. Available at: http://CRAN.R-project.org/package=vegan.

Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289-290.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & the R Development Core Team (2011) Nlme: linear and nonlinear mixed effects models. R package version 3.1-102. Available at: http://cran.r-project.org/web/packages/nlme/index.html.

Price, T.D., Helbig, A.J. & Richman, A.D. (1997) Evolution of breeding distributions in the old world leaf warblers (genus *Phylloscopus*). *Evolution*, **51**, 552–561.

Quintero, I. & Wiens, J.J. (2013) What determines the climatic niche width of species? The role of spatial and temporal climatic variation in three vertebrate clades. *Global Ecology and Biogeography*, **22**, 422–432.

R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org.

Räsänen, K. & Hendry, A.P. (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters*, **11**, 624–636.

Rex, M.A., Stuart, C.T., Hessler, R.R., Allen, J.A., Sanders, H.L. & Wilson, G.D.F. (1993) Global-scale latitudinal patterns of species diversity in the deep-sea benthos. *Nature*, **165**, 636–639.

Rohde, K. (1996) Rapoport's rule is a local phenomenon and cannot explain latitudinal gradients in species diversity. *Biodiversity Letters*, **3**, 10–13.

Ruffo, S. & Stoch, F. (2006) Checklist and distribution of the Italian fauna. 10,000 terrestrial and inland waters species. *Memorie del Museo Civico di Storia Naturale di Verona, 2.serie, Sezione Scienze della Vita*, **17**, 1–307 (data on CD-ROM)

Ruggiero, A. & Werenkraut, V. (2007) One-dimensional analyses of Rapoport's rule reviewed through meta-analysis. *Global Ecology and Biogeography*, **16**, 401–414.

Sastre, P., Roca, P., Lobo, J.M. & EDIT co-workers (2009) A Geoplatform for the accessibility to environmental cartography. *Journal of Biogeography*, **36**, 568.

Soberón, J., Jimenéz, R., Golubov, J., Koleff, P. (2007) Assessing completeness of biodiversity databases at different spatial scales. *Ecography*, **30**, 152–160.

Stein, H., Griebler, C., Berkhoff, S., Matzke, D., Fuchs, A. & Hahn, H.J. (2012) Stygoregions – a promising approach to a bioregional classification of groundwater systems. *Scientific. Reports*, **2**, 1–9.

Stevens, G.C. (1989) The latitudinal gradient in geographical range – how so many species co-exist in the tropics. *American Naturalist*, **133**, 240–256.

Svenning, J.C., Fløjgaard, C. & Baselga, A. (2011) Climate, history and neutrality as drivers of mammal beta diversity in Europe: insights from multiscale deconstruction. *Journal of Animal Ecology*, **80**, 393–402.

Trabucco, A. & Zomer, R.J. (2010) Global soil water balance geospatial database. CGIAR Consortium for Spatial Information. Published online, available from the CGIAR-CSI GeoPortal at: http://www.cgiar-csi.org.

Trontelj, P., Douady, C.J., Fišer, C., Gibert, J., Gorički, S., Lefébure, T., Sket, B. & Zakšek, V. (2009) A molecular test for cryptic diversity in ground water: how large are the ranges of macro-stygobionts? *Freshwater Biology*, **54**, 727–744.

Veter, N.M., DeSantis, L.R.G., Yann, L.T., Donohue, S.L., Haupt, R.J., Corapi, S.E., Fathel, S.L., Gootee, E.K., Loffredo L.F., Romer J.L. & Velkovsky S.M. (2013) Is Rapoport's rule a recent phenomenon ? A deep time perspective on potential causal mechanisms. *Biology Letters*, **9**, 20130398.

Vogt, J., Soille, P., de Jager, A., Rimavičiūtė, E., Mehl, W., Foisneau, S., Bódis, K., Dusart, J., Paracchini, M.L., Haastrup, P. & Bamps, C. (2007): A pan-European River and Catchment Database. European Commission – Joint Research Centre, Luxembourg, (EUR 22920 EN), 120 pp.

Weiser, M.D., Enquist, B.J., Boyle, B., Killeen, T.J., Jørgensen, P.M., Fonseca, G., Jennings, M.D., Kerkhoff, A.J., Lacher Jr., T.E., Monteagudo, A., Núñez Vargas, M.P., Phillips, O.L., Swenson, N.G. & Vásquez Martínez, R. (2007) Latitudinal patterns of range size and species richness of New World woody plants. *Global Ecology and Biogeography*, **16**, 679–688.

Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, **73**, 3–36.

Zagmajster, M., Culver, D.C., Christman, M.C. & Sket, B. (2010) Evaluating the sampling bias in pattern of subterranean species richness: combining approaches. *Biodiversity and Conservation*, **19**, 3035–3048.

BIOSKETCHES

Maja Zagmajster is a researcher in Subterranean Biology Laboratory at Department of Biology at University of Ljubljana, Slovenia. Her research interests focus on biogeography, biodiversity, distribution patterns and ecology of subterranean fauna.

David Eme is a doctoral student in the team "Ecology, Evolution and Ecosystem Studies in groundwater" at the University of Lyon, France. He is using statistical approaches in biogeography and phylogeography to reveal historical and contemporary determinants of biodiversity patterns.

The EGCD initiative was launched by F.M. in 2010. It gathers ecologists with expertise in the taxonomy of groundwater taxa including Copepoda (D.G. and F.S), Podocopida (P.M.), Isopoda (F.M. and D.E.), and Amphipoda (C.F. and M.Z.), geographical information systems (J.-F.C.) and data analysis (D.E.). All authors contributed significantly to the conception of this study and final version of the manuscript.

SUPPORTING INFORMATION

Appendix S1: Relationships between median range size or species richness and latitude.
Appendix S2: Latitudinal patterns of area of occupancy.
Appendix S3: Relationships between median range size or species richness and latitude, while accounting for sampling bias.
Appendix S4: Models for determining drivers of spatial variation in median range size.

2.3) Article 3 Multi-causality and spatial non-stationarity in the determinants of groundwater crustacean diversity in Europe

David Eme[1], Maja Zagmajster[2], Cene Fišer[2], Diana Galassi[3], Pierre Marmonier[1], Fabio Stoch[3], Jean-Francois Cornu[4], Thierry Oberdorff[5] and Florian Malard[1]

[1]Université de Lyon, UMR5023 Ecologie des hydrosystèmes Naturels et Anthropisés, Université Lyon1, ENTPE, CNRS, 6 rue Raphaël Dubois, 69622 Villeurbanne, France.

[2]SubBioLab, Department of Biology, Biotechnical Faculty, University of Ljubljana, Večna pot 111, p.p. 2995, SI-1000 Ljubljana, Slovenia.

[3]Department of Life, Health & Environmental Sciences, University of L'Aquila, Via Vetoio, Coppito, I-67100, L'Aquila, Italy.

[4]Institut des Sciences de l'Evolution (ISEM), CNRS-IRD-UM2, Université de Montpellier 2, 34095 Montpellier Cedex, France.

[5]UMR BOREA, Département Milieux et Peuplement Aquatiques, MNHN, CNRS 7208, IRD 207, UPMC, Muséum National d'Histoire Naturelle, 43, rue Cuvier ; 75231 Paris, France.

Soumis à *Ecography*

Abstract

The recognition of multi-causality and spatial non-stationarity in the determinants of large-scale biodiversity patterns requires to consider the role of multiple mechanisms, their interactions, and how these mechanisms vary in strength relative to each other across geographical space. Here, we challenge the view that historical climate stability primarily drives European patterns of groundwater crustacean diversity by testing also the role of spatial heterogeneity and productive energy. We predicted that the three mechanisms would be equally important at continental scale, but that they would exhibit predictable latitudinal changes in their relative strength. To test this prediction, we selected predictors representing each mechanism and analyzed separately and jointly their effects and interactions using global regression models. We further mapped the independent and overlapping effects of mechanisms across Europe using partial geographically weighted regressions. When analyzed separately, the three mechanisms explained the same amount of variation in species richness, but in the joint analysis, the influence of historical climate stability became hidden in the variation shared with the other mechanisms. Topographic heterogeneity interacted synergistically with actual evapotranspiration and habitat heterogeneity on species richness. Spatial non-stationarity in the independent and overlapping effects of the three mechanisms was the most plausible explanation for the hump-shaped latitudinal pattern of crustacean species richness. Productive energy and spatial heterogeneity were important predictors at mid and southern latitudes, whereas historical climate stability overlapped with the two other mechanisms and productive energy in northern and southern Europe, respectively. Our finding underlines the danger of looking for evidence to support the role of one mechanism without testing whether that evidence may also support the effect of other mechanisms. Multi-causality and spatial non-stationarity provide a broader perspective of groundwater

biodiversity determinants that revives the importance of spatial heterogeneity and the strong dependence of subterranean communities on food supply from the surface.

Introduction

A major goal in biogeography and macroecology is to understand the causes of taxonomic diversity gradients (Brown 1995, Field et al. 2009). The number of species in a region is ultimately determined by three processes - speciation, extinction and dispersal (Brown 1995, Wiens 2011) - but identifying the main factors acting upon these processes is still a matter of intense debate (Whittaker et al. 2001, Currie et al. 2004, Field et al. 2009). A multitude of hypotheses were proposed to explain geographic variation of species richness at large spatial scales (Palmer 1994, Willig et al. 2003). Yet, in a desire of synthesis, all of them have been progressively merged into three broad mechanisms: spatial heterogeneity, climate/productivity, and history (Field et al. 2009, Oberdorff et al. 2011, Tisseuil et al. 2013). Higher habitat heterogeneity can inflate species richness by increasing speciation through increased specialization and (or) by promoting local survival (Kerr and Packer 1997, Rahbek and Graves 2001, Davies et al. 2007, Ohlemüller et al. 2008). Larger areas tend to contain more species (Brown 1995) and geometric constraints have been proposed to explain peaks of species richness in centre of bounded domains (i.e. the mid-domain effect [MDE]; Colwell et al. 2009). The climate/productivity hypothesis proposes that climate influences the number of species either directly through physiological effects, or indirectly by controlling resource productivity (Brown et al. 2004, Currie et al. 2004, Evans et al. 2005, Hawkins et al. 2003, Field et al. 2009). The historical hypothesis associates the long-lasting effect of past climatic events on species richness to dispersal constraints (Oberdorff et al. 2011). Species extinction has supposedly been more severe in regions that experienced large Quaternary climatic oscillations (Dynesius and Jansson 2000, Araújo et al. 2008, Leprieur et al. 2011). Yet, the fingerprint of past climate change on present-day pattern of species richness depends on the differential ability of taxa to colonize vacant habitats (Hof et al. 2008).

Multi-causality, rather than the role of a single mechanism, is now recognized as the most plausible explanation of species richness patterns at continental to global scales (Whittaker et al. 2001, Hawkins et al. 2003, Brown et al. 2004, Field et al. 2009, Tisseuil et al. 2013, Gouveia et al. 2013). Yet, the relative influence of the three broad mechanisms has been tested in a relatively small number of taxa, essentially vascular plants and vertebrates (Beck et al. 2012). Moreover, their influence has generally been evaluated using global regression models that most often ignore spatial non-stationarity in the relationship between richness and environmental predictors (Foody 2004, Eiserhardt et al. 2011; but see Kerr and Packer 1997, Hawkins et al. 2003). In complement to these global models, local regression models can reveal singularities in the mechanisms shaping species richness pattern among regions (Brundson et al. 1996, Foody 2004, Svenning et al. 2009, Eiserhardt et al. 2011, Gouveia et al. 2013). Another obstacle to disentangling the role of distinct mechanisms is collinearity between predictors. For example, temperature seasonality often covaries with long-term climate variability and productivity (Gouveia et al. 2013, Morueta-Holme et al. 2013). Yet, this obstacle can be overcome if the studied taxa are not exposed to spatial variation in temperature seasonality.

Groundwater habitats (i.e. all temporally and permanently water-saturated zones in the subsurface) can offer useful case studies to tease apart mechanisms driving patterns of species richness because they show little intra-annual temperature variation. Therefore, the effect of long–term temperature oscillations cannot be masked by co-variation with temperature seasonality (Zagmajster et al. 2014). Moreover, the role of spatial heterogeneity, contemporary climate and history is not going to

be obscured by a strong-mid domain effect (Colwell et al. 2004) because most groundwater species have extent of occurrence typically less than 200 km (Trontelj et al. 2009, Deharveng et al. 2009). Coarse-grain studies of subterranean species richness pattern in Europe documented a northward monotonic decline in the number of species per biogeographic regions or countries (Hof et al. 2008, Stoch and Galassi 2010). This pattern has long been solely attributed to the disproportionate extinction of small range species at higher latitudes during cold Pleistocene climate coupled with limited postglacial colonization due the weak propensity for dispersal among groundwater organisms (Hof et al. 2008). More recently, Cornu et al. (2013) brought evidence that lower habitat heterogeneity might also be responsible, in addition to historic climate variability, for the lower species richness of groundwater communities in northern Europe. Finer-grain studies revealed a somewhat different latitudinal biodiversity pattern with a ridge of high species richness at latitudes ranging from ca 42 to 46° N, in both the terrestrial and aquatic subterranean fauna (Culver et al. 2006, Zagmajster et al. 2014). This hump-shaped latitudinal pattern of species richness appeared to be robust to sampling bias, at least for the aquatic subterranean fauna (Zagmajster et al. 2014). Culver and co-authors (2006) suggested that the ridge of high biodiversity in the terrestrial cave fauna was associated with regions of high surface productivity and cave density. The productive energy hypothesis reflects the strong dependence of subterranean communities on food supply from the surface because there is no primary production from photosynthesis in the subsurface (Gibert and Deharveng 2002). Yet, this broader perspective of groundwater biodiversity patterns involving multi-causality in the explanation of species richness patterns and spatial non-stationarity of the causal mechanisms awaits rigorous testing (Griebler et al. 2014).

The present study provides the first quantitative assessment of the relative importance of historical climate stability, productive energy and spatial heterogeneity in shaping species richness patterns of groundwater crustaceans at a continental scale. First, we predicted that these broad mechanisms would be equally important in explaining geographic variation in species richness in Europe. To test this prediction, we selected a set of predictors representing the three mechanisms and analyzed separately and jointly their effects and interactions using global regression models. Second, we predicted that the effect of productive energy and spatial heterogeneity would be stronger at latitudes corresponding to the ridge of high species richness, whereas the role of historical climate stability would gain in importance further north of the ridge. To test this second prediction, we used partial geographically weighted regression (GWR) models to partition and map the independent and shared effects of the three mechanisms across space in Europe.

Material and Methods

Species richness data set

We used the European groundwater crustacean data set (EGCD) recently published by Zagmajster et al. (2014). This is the most comprehensive occurrence data set currently available at the European scale, with a total of 21,700 occurrences for 1570 species and subspecies of obligate groundwater crustaceans. Crustaceans are the most diversified taxa in groundwater, representing more than 65% of all obligate groundwater species presently known from Europe (Deharveng et al. 2009, Stoch and Galassi 2010). Occurrence data are restricted to species that complete their entire life cycle exclusively in groundwater, but they are from a variety of habitats in consolidated rock aquifers (e.g. vadose and saturated zones of karst aquifers), unconsolidated sediment aquifers (e.g. phreatic zone of alluvial aquifer) and ecotonal zones between groundwater and surface water (e.g. the hyporheic zone of streams). Species occurrences were projected onto the grid cell system of 0.9 × 0.9° spatial

resolution provided by the EDIT geoplatform (Sastre et al. 2009). Cell area was kept constant (10,000 km^2) all over the grid by adjusting the longitudinal divisions between adjacent cells in each latitudinal band. Coastal cells with < 20% of land area were excluded from the grid (except for islands). This yielded a total of 701 cells, among which 494 contained at least one species occurrence. Richness was defined as the number of species and subspecies contained in each cell because subspecies were considered as distinct phylogenetic units (Zagmajster et al. 2014).

Predictors of species richness

For each grid cell, we quantified five predictors in order to test the three broad mechanisms on the drivers of species richness. To measure historical climate stability, we used temperature and precipitation anomalies, defined as the differences in mean annual temperature and annual precipitation between the present and LGM (Araújo et al. 2008, Leprieur et al. 2011, Gouveia et al. 2013). Temperature and precipitation during the LGM were obtained from two global circulation models, namely CCSM and MIROC2 (Hijmans et al. 2005; data available at http://www.worldclim.org/) and temperature and precipitation anomalies from both were averaged to account for variation between models. To test for the energy hypothesis, we initially used mean actual evapotranspiration (AET), mean annual air temperature, and mean annual precipitation as surrogates of productive energy, ambient energy and water availability, respectively (Evans et al. 2005). However, we only retained AET because mean annual temperature and mean annual precipitation were collinear with temperature anomaly and elevation range, respectively (variance inflation factor [VIF] > 5; Zuur et al. 2010). AET provides a synthetic index of water–energy dynamics (O'Brien 2006), which subsumes ambient energy and water availability, two crucial factors determining the amount of plant productivity (Whittaker et al. 2001, Evans et al. 2005). Data on AET were extracted from the 30 arc-second resolution world map released by Trabucco and Zomer (2010). To represent spatial heterogeneity, we calculated elevation range and groundwater habitat diversity. Elevation range is a surrogate of topographic heterogeneity (Rahbek and Graves 2001, Davies et al. 2007, Leprieur et al. 2011) and was estimated as the highest difference in elevation between any two locations in a cell using elevation data from the pan-European River and catchment database (Voght et al. 2007). Groundwater habitat diversity was estimated for each cell using the Shannon's diversity index defined as $H' = - \Sigma\ pi \times \ln pi$, where pi represented the areal proportion of 12 groundwater habitat types (excluding non aquiferous rocks). Habitat types were distinguished according to Cornu et al. (2013) based on the permeability of the rock, the type of voids (i.e. pore and fissures) and their size. The areal proportion of each habitat was computed using the vector map of groundwater habitats in Europe (available at: http://data.freshwaterbiodiversity.eu/data/shapefiles/). Elevation range and temperature anomaly were ln-transformed to satisfy normality assumption and all variables were standardized. We checked for multi-collinearity among predictors using variance inflation factors and found them to be in acceptable range (VIF < 5; Zuur et al. 2010).

Statistical analysis

Global models

To test the first prediction that the three broad mechanisms would be equally important in shaping species richness patterns, we performed ordinary least square (OLS) models using ln transformed species richness as a response variable to improve variance homogeneity. All models were fitted to the 494 grid cells containing at least one species because we could not distinguish zero richness values from sampling gaps. To allow for the possibility of minor nonlinearity in the

relationship between species richness and the predictors, the later were included in the models in their linear and quadratic forms. We used a three-step procedure to test for multi-causality. In step 1, the amount of variance explained by each mechanism was assessed separately. To evaluate the relative importance of predictors within each broad mechanism (including their linear and quadratic terms), we used multi-model inferences based on information theory (Burnham and Anderson 2002). We ran all possible OLS models and retained only those models whose difference in the Akaike's information criterion corrected for small sample size (AICc) with the best model (lowest AICc) was ≤ 5. Using this model subset, we measured the relative importance of each predictor as the sum of AICc weights of models in which the predictor occurred (Burnham and Anderson 2002). The amount of explained variance of the best model was estimated using the adjusted R^2. In step 2, the amount of variance explained by the three mechanisms was assessed jointly in an additive model using the best set of predictors selected separately for each mechanism (see above). Then, we performed variance partitioning (Legendre and Legendre 1998) to estimate the independent and shared contributions of the three mechanisms. In step 3, we included interaction terms between predictors in the joint model and tested the improvement of the model fit over the joint additive model using AICc and explained variance. For both joint models (additive and with interactions), we used multi-model inferences and AICc weights to evaluate the relative importance of predictors and interaction terms.

Global models with spatial autocorrelation

We re-ran steps 1 to 3 of the analysis using simultaneous autoregressive (SAR) models (Kissling and Carl 2007), because spatial autocorrelation in the residuals of OLS models could potentially affect the estimation of parameters and their statistical significance (Davies et al. 2007, Dormann et al. 2007). To select the most appropriate SAR models (i.e. the one with no spatial autocorrelation in the residuals and minimum AICc), we tested a range of neighborhood distances between cells (i.e. 220-800 km) using a row standardized coding scheme for the spatial weight matrix. The total variance explained by the best SAR model was computed using pseudo R^2 (including the spatial component), whereas variance partitioning of the three broad mechanisms was calculated using partial-pseudo R^2 values (excluding the spatial component; Araújo et al. 2008, Morueta-Holme et al. 2013).

All analyses were run in R statistical software (R core Development Team, 2013). Model selection and multi-model inference for OLS and SAR were performed using MuMIn R package (Barton, 2013). Neighborhood distance matrices and SAR models were computed with the spdep R package (Bivand et al. 2012). Variation partitioning was computed with the vegan R package (Oksanen et al. 2012) for OLS and with a simple set of equations following Legendre and Legendre (1998) for SAR models.

Local model

We used GWR models (Brundson et al. 1996) to test the second prediction of a change in the relative influence of the three broad mechanisms among regions in Europe. GWR estimates independent regression parameters (coefficients, errors and R^2) for each grid cell, wherein all cells within a given bandwidth are weighted according to their geographical distance to the focal grid cell using a predefined spatial kernel function (Brundson et al. 1996). Herein, we used a Gaussian function for the decay of the weight with distance. The use of a fixed band width may cause spurious inferences in GWR models if the number of neighbors among focal cells is too variable (Svenning et al. 2009). To avoid this problem, we fixed the number of neighbors to a given percentage of all cells in the grid. To ensure a local modeling with sufficient neighboring cells per local regression, the percentage of neighbors was optimized (minimum AICc) by searching between 7 and 30% of all cells with the

Golden Section Search (Svenning et al. 2009). The five predictors were integrated in a full GWR model, but quadratic terms were discarded because GWR is particularly sensitive to multi-collinearity among predictors (Wheeler and Tiefelsdorf 2005). To map the independent and shared effects of the three mechanisms across space in Europe, we performed a cell-by-cell variance partitioning using partial GWR models (see Eiserhardt et al. 2011, Gouveia et al. 2013 for a similar application). GWR and partial GWR models were computed using SAM v4.0 (Rangel et al. 2010).

Results

Global models

When evaluating separately, the amount of variance explained by historical climate stability, productive energy and spatial heterogeneity was roughly similar, regardless of the models used (adjusted R^2 = 0.345 – 0.393, Table 1). For each mechanism, taking into account spatial autocorrelation greatly improved the model fit. Species richness showed a hump-shaped relationship with temperature anomaly whereas it was positively related to AET, elevation range and habitat diversity (Supplementary material Appendix 1, see parameter estimates in Table A1 and response curves in Fig. A1).

The joint additive model explained more than half of the variance in species richness (adjusted R^2 = 0.518 and SAR pseudo R^2 = 0.624, Table 2, Supplementary material Appendix 2, see parameter estimates in Table A2). Variance partitioning attributed a larger proportion of variance to the independent effects of productive energy (9%) and spatial heterogeneity (5.8%) than to the effect of historical climate stability (0.4%) (Table 2). A large fraction of variance was shared between the three mechanisms (18 to 21%) and between productive energy and historical climate stability (7.9 to 9.3%). We found little difference in the amount of variance attributed to the independent effects of the three mechanisms between OLS and SAR models (Table 2). Yet, taking into account spatial autocorrelation increased the shared component of productive energy and spatial heterogeneity (from 1.6 to 5.9%) and decreased the shared component of historical climate stability and spatial heterogeneity (from 6.2 to 2.1%). Comparison of summed AICc weights among predictors showed that temperature and precipitation anomalies were less important predictors relative to AET, elevation range and groundwater habitat diversity (Table 3).

Joint models with interactions explained slightly more variance than joint additive models and had lower AICc scores, indicating that interactions improved the model fit (Table 3, Supplementary material Appendix 2, see parameter estimates in Table A2). Elevation range interacted positively with AET and groundwater habitat diversity (Figure 2, Fig. A2). The summed AICc weights for both interactions in the OLS model was 1, indicating that they were as important as the effects of single predictors in explaining geographic variation in species richness. Incorporating interaction terms in

Table 1: Summary results of the best models for testing separately the three broad mechanisms on the drivers of species richness of obligate groundwater crustaceans in Europe. ExVar: proportion of explained variance (%); AICc: Akaike's information criterion for small sample size; Ano_T: temperature anomaly; Ano_P: precipitation anomaly; AET: actual evapotranspiration; Elevr: elevation range; Hab: habitat diversity. The superscript next to the predictor's names indicates the quadratic form. For simultaneous autoregressive models, the best neighborhood distance was 300 km.

Model	Historical climate stability						Productive energy				Spatial heterogeneity					
			AICc weights						AICc weights				AICc weights			
	ExVar	AICc	Ano_T	Ano_T^2	Ano_P	Ano_P^2	ExVar	AICc	AET	AET^2	ExVar	AICc	Elevr	Elevr^2	Hab	Hab^2
Ordinary least square models	35.4	1158	1	1	0.71	0.43 [a]	39.5	1126	1	0.67	34.46	1165	1	1	1	0.26 [a]
Simultaneous autoregressive models	58.02	998.1	0.68	1	0.59	1	59.65	970.2	1	0.32 [a]	59.28	986.3	1	0.33 [a]	1	0.47 [a]

[a] Terms excluded in the joint models (see Table 3)

Table 2: Independent and shared contributions of historical climate stability (H), productive energy (E) and spatial heterogeneity (S) to variation in species richness of obligate groundwater crustaceans in Europe. In abbreviations, colons denote shared variance between mechanisms. For simultaneous autoregressive models, the best neighborhood distance was 220 km.

Model	Explained variance (%)	Variance partitioning (%)						
		Historical climate stability (H)	Productive energy (E)	Spatial heterogeneity (S)	H:E	H:S	E:S	H:E:S
Ordinary least square models	51.77	0.44	9.01	5.83	7.87	6.23	1.57	20.83
Simultaneous autoregressive models	50.82 (62.43) [a]	0.09	9.27	5.74	9.29	2.06	5.89	18.47

[a] including the spatial component

OLS and SAR models increased the summed AICc weight of temperature anomaly but it did not downweight the importance of AET, elevation range and habitat diversity (Table 3).

Local model

The GWR model explained 61.5% of total variance in species richness (model AICc = 941.3), but local adjusted R^2 varied considerably across space (local adj.R^2 = 0.10 – 0.78). Local adjusted R^2 values were substantially higher in western than in eastern regions (see Fig. 1h). The latitudinal ridge of high species richness was recovered by the model but predicted values of species richness in the ridge were lower than observed values, more particularly in Eastern Europe (Fig. 1g, see also Supplementary material Appendix 3, Fig. A3).

Cell-by-cell variance partitioning revealed substantial geographic variation in the relative proportion of explained variance attributed to the independent and shared effects of the three broad mechanisms (Fig. 3). Spatial heterogeneity was an important predictor of groundwater crustacean species richness in south-western Europe whereas the independent effect of productive energy was substantial in the southern margins of Central Alps and the periphery of the Iberian Peninsula. Contrary to our second prediction, the independent effect of historical climate variability was consistently low, even in northern Europe (Fig. 3). Yet, a striking result of partial GWR model was the substantial amount of variance shared by the three mechanisms in a vast region of Europe extending from 48 and 62° north latitude (Fig. 3). This indicated that the northward decline of species richness reflected the joint effects of a higher temperature anomaly and lower productive energy and spatial heterogeneity. Finally, the effect of historical climate stability was also inseparable from that of productive energy in central Spain.

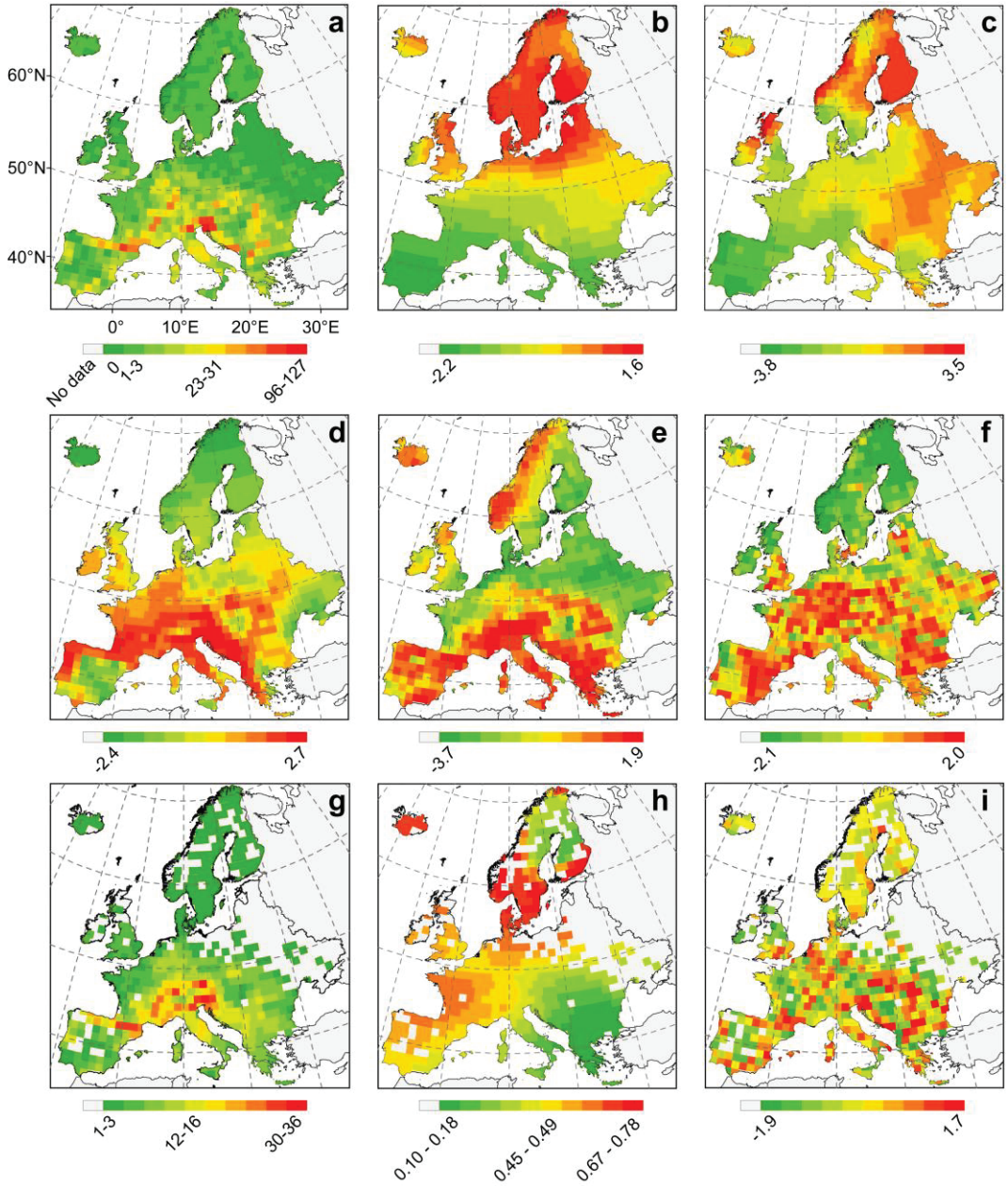


Figure 1: Top and middle rows: patterns of species richness and environmental predictors across Europe; a) species richness of obligate groundwater crustaceans; b) temperature anomaly (range: 1.8-29 °C); c) precipitation anomaly (range: -661-625 mm), d) actual evapotranspiration (range: 264-764 $mm.yr^{-1}$), e) elevation range (range: 7-3733 m), f) habitat diversity (range: 0-2). Color scale for environmental predictors corresponds to standardized values. Bottom row: geographically weighted regression (GWR) results for species richness; g) estimated richness for cells containing at least one species (range: 1-36); h) adjusted R^2; i) model residuals.

Table 3: Summary results of the best models for testing jointly the three broad mechanisms on the drivers of species richness of obligate groundwater crustaceans in Europe. OLS: ordinary least square models; SAR: simultaneous autoregressive models; ExVar: proportion of explained variance (%); AICc: Akaike's information criterion for small sample size; Ano_T: temperature anomaly; Ano_P: precipitation anomaly; AET: actual evapotranspiration; Elevr: elevation range; Hab: habitat diversity. The superscript next to the predictor's names indicates the quadratic form. For SAR, the best neighborhood distance was 220 km.

Model		ExVar	AICc	Summed AICc weights									Interactions
				Historical climate stability				Productive energy		Spatial heterogeneity			
				Ano_T	Ano_T^2	Ano_P	Ano_P^2	AET	AET^2	Elevr	Elevr^2	Hab	
Additive	OLS	51.77	1019	0.87	0.61	0.29	- [b]	1	0.79	1	0.97	1	na [c]
	SAR	50.82 (62.43) [a]	936	0.48	0.56	0.36	0.48	1	-	1	-	1	na
With interactions	OLS	55.05	984	1	1	-	-	1	0.67	1	0.23	1	Elevr × AET = 1; Elevr × Hab = 1
	SAR	54.47 (63.18)	918	0.80	0.87	0.46	0.37	1	-	1	-	1	Elevr × AET = 1; Elevr × Hab = 0.50

[a] including the spatial component

[b] not included in the model selection

[c] na: not applicable

Discussion

Our first prediction that historical climate stability, productive energy and spatial heterogeneity would be equally important in explaining geographic variation in species richness in Europe was supported when the importance of each mechanism was assessed separately. However, it was no longer supported when the three mechanisms were analyzed together in a joint additive model because the independent effect of historical climate stability was marginal relative to productive energy and spatial heterogeneity. The whole influence of historical climate stability became hidden in the variation shared with the two other mechanisms (but see below for a discussion of shared variance). Our finding underlines the danger of looking for evidence to support a particular mechanism without testing whether that evidence may be confounded by the effects of other mechanisms.

This study provides the first evidence that AET is an important predictor of groundwater species richness patterns at continental scale, thereby supporting the hypothesis that differences in the number of species among regions in part reflect the strong dependence of subterranean ecosystems on surface productivity (Culver et al. 2006). Water-energy dynamics is the most important driver of large-scale variation in richness across a wide range of plant and animal groups (Hawkins et al. 2003, Field et al. 2009). However, the importance of food supply to groundwater had until now been restricted to the explanation of differences in the number of coexisting species between local communities (Datry et al. 2005, Foulquier et al. 2011). The relationship between productive energy and groundwater species richness can be hump-shaped at local scale (Strayer et al. 1997). In the absence of oxygen production in groundwater, species richness increases with increasing organic matter supply until dissolved oxygen deficiency resulting from microbial respiration becomes a limiting factor. At continental scale, we found that the richness of groundwater crustaceans increased monotonically with AET. Yet, the explanatory power of AET when modeled separately (adjusted $R^2 = 0.4$, Fig. S1) was lower than that observed for a wide range of taxa and habitats (see Hawkins et al. 2003, Field et al. 2009), most probably because AET is a distal surrogate of energy supply to groundwater. Indeed, the amount of organic matter reaching the groundwater table not only depends on organic matter production in the surface environment but also on water infiltration rate and significance of retention and degradation processes in the infiltration zone (Datry et al. 2005, Foulquier et al. 2010). However, the latter two parameters are not available for mapping at a European scale.

Spatial heterogeneity was also an important correlate of groundwater crustacean diversity as both elevation range and habitat diversity were positively related to species richness. Strong topographic heterogeneity imposes barriers to dispersal and generates steep ecological gradients, which both contribute to reduce gene flow among populations and can ultimately lead to speciation (Qian and Ricklefs 2000, Rahbek and Graves 2001, Dias et al. 2013). Sharp climatic gradients in regions of high topographic heterogeneity can also decrease the extinction rate of species with poor dispersal capacity by increasing the probability that they survive changing climate in nearby refugia (Ohlemüller et al. 2008). Eme et al. (2013) recently brought molecular evidence that the Jura and Alpine foothills acted both as diversification hotspots and Pleistocene refugia among groundwater isopods of the genus *Proasellus*. Habitat heterogeneity is another important predictor of species richness but there is still much debate about whether the relationship is positive or unimodal (Hortal et al. 2013). Here, we found that the richness of groundwater crustaceans increased monotonically with habitat heterogeneity as represented by the number and relative proportion of distinct habitat types with characteristic flow conditions (permeability) and pore size. Trontelj et al. (2012) found clear evidence of divergent morphological adaptations among closely related species of amphipods that use different cave microhabitats with distinct flow velocity (but see also Fišer et al. 2012). A similar

process of divergent selection may in part account for a positive richness - habitat heterogeneity relationship at a continental scale, because it allows more species to co-exist in regions characterized by a high diversity of aquifers in unconsolidated sediments and consolidated rock (Cornu et al. 2013).

Testing for multi-causality in the explanation of groundwater crustacean species richness in Europe also requires to better understand the role of interactions between predictors within and among mechanisms. Our finding that the positive relationship between elevation range and species richness becomes steeper in regions of high AET supports a synergistic effect between energy and spatial heterogeneity (Kerr and Packer 1997, Rahbek and Graves 2001, Ruggiero and Kitzberger 2004). This may indicate that the potential for spatial heterogeneity to increase speciation or decrease extinction becomes stronger when productive energy is high enough to maintain large population size (Evans et al. 2005). The positive interaction between elevation range and habitat diversity highlights another synergetic effect on species richness between two components of spatial heterogeneity that have rarely been separated in macroecology (Kerr et al. 2001). It suggests that allopatric speciation caused by topographic highs may add up to niche-based mechanisms operating within a habitat-rich landscape.

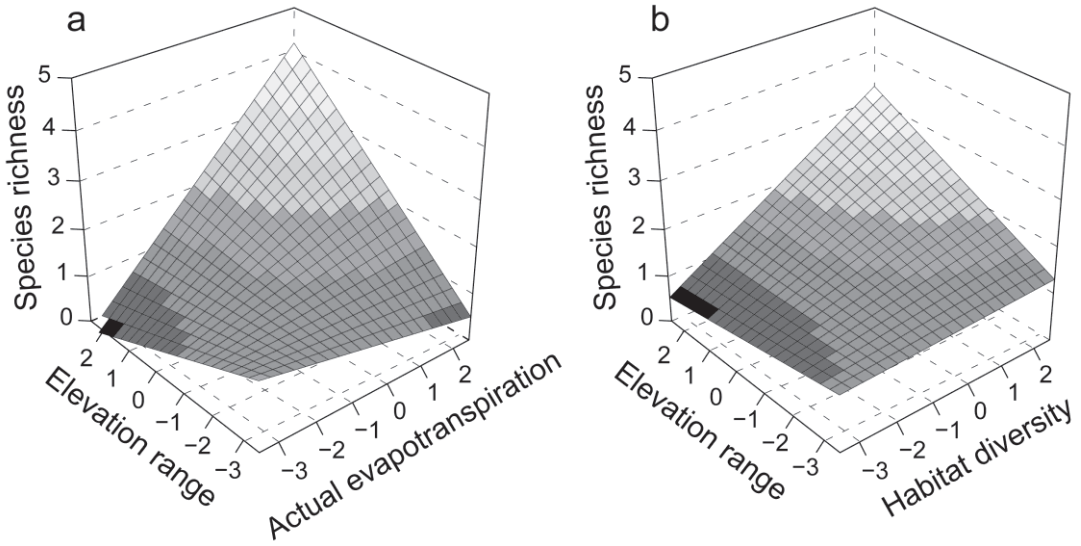


Figure 2: Interaction effects between elevation range and actual evapotranspiration (a) and habitat diversity (b) on species richness of obligate groundwater crustaceans. Species richness values are fitted values from simultaneous autoregressive models. Species richness values are ln transformed and scale for environmental predictors corresponds to standardized values.

We found clear evidence of spatial non-stationarity in the relative importance of the three mechanisms but we also identified regions where a large amount of variation could not be either explained or separated among mechanisms. Yet, our second prediction that the independent effects of productive energy and spatial heterogeneity would be stronger along the ridge of high species richness, whereas the independent effect of historical climate stability would gain in importance north of the ridge, was not entirely supported by the outputs of partial GWR-models. The independent effect of spatial heterogeneity prevailed in south-western Europe and productive energy accounted for a substantial amount of variation in species richness not only in regions located along the ridge (e.g. Cantabria, Pyrenees, south of the Italian and Slovenia Alps) but also at the periphery of the Iberian Peninsula. Moreover, the independent effect of historical climate stability was low all over Europe, although the latter mechanism covaried with productive energy and spatial heterogeneity in northern Europe and with productive energy in the Iberian Peninsula. The substantial amount of variance shared by the three mechanisms at latitudes ranging from 48 to 62° N suggests that their joint effects

rather than the individual impact of historical climate stability produce the sharp decrease in groundwater crustacean species richness at higher latitudes. The historical climate hypothesis suggests that many groundwater species must have gone extinct in northern European regions affected by cold Pleistocene climates because small range species are particularly prone to extinction (Dynesius and Jansson 2000). Yet, postglacial colonization of vacant habitats may be further restricted if the mismatch between low food supply and increased energetic demand due to dispersal caused reduced fitness among small-bodied species that can colonize northern European aquifers characterized by small pore size and reduced permeability (Cornu et al. 2013). Moreover, low food availability in groundwater tends to favor omnivores (Gibert and Deharveng 2002), which may monopolize food resources and prevent the establishment of late colonizers (Urban et al. 2008). Historical climate stability also co-varied with productive energy in the Iberian Peninsula. Since the LGM, the inner Iberian plateau (i.e. Meseta Central) has become comparatively more arid than the periphery of the Iberian Peninsula which has higher AET. The trend over time of increasing drought, combined with lower productive energy, might have increased extinction rates, thereby resulting in lower species richness in Central Spain. Despite the impossibility of isolating the independent effect of historical climates, our results suggest that the long lasting effect of cold Pleistocene climates in northern Europe and increasing aridity in the Iberian Peninsula might partly explain the hump-shaped latitudinal pattern of groundwater crustacean species richness in Europe. This follows Culver and co-authors' suggestion that the mid-latitude ridge in terrestrial cave fauna might correspond to regions where, in the absence of dry or cold events, productivity remained high over recent geological time (Culver et al. 2006).

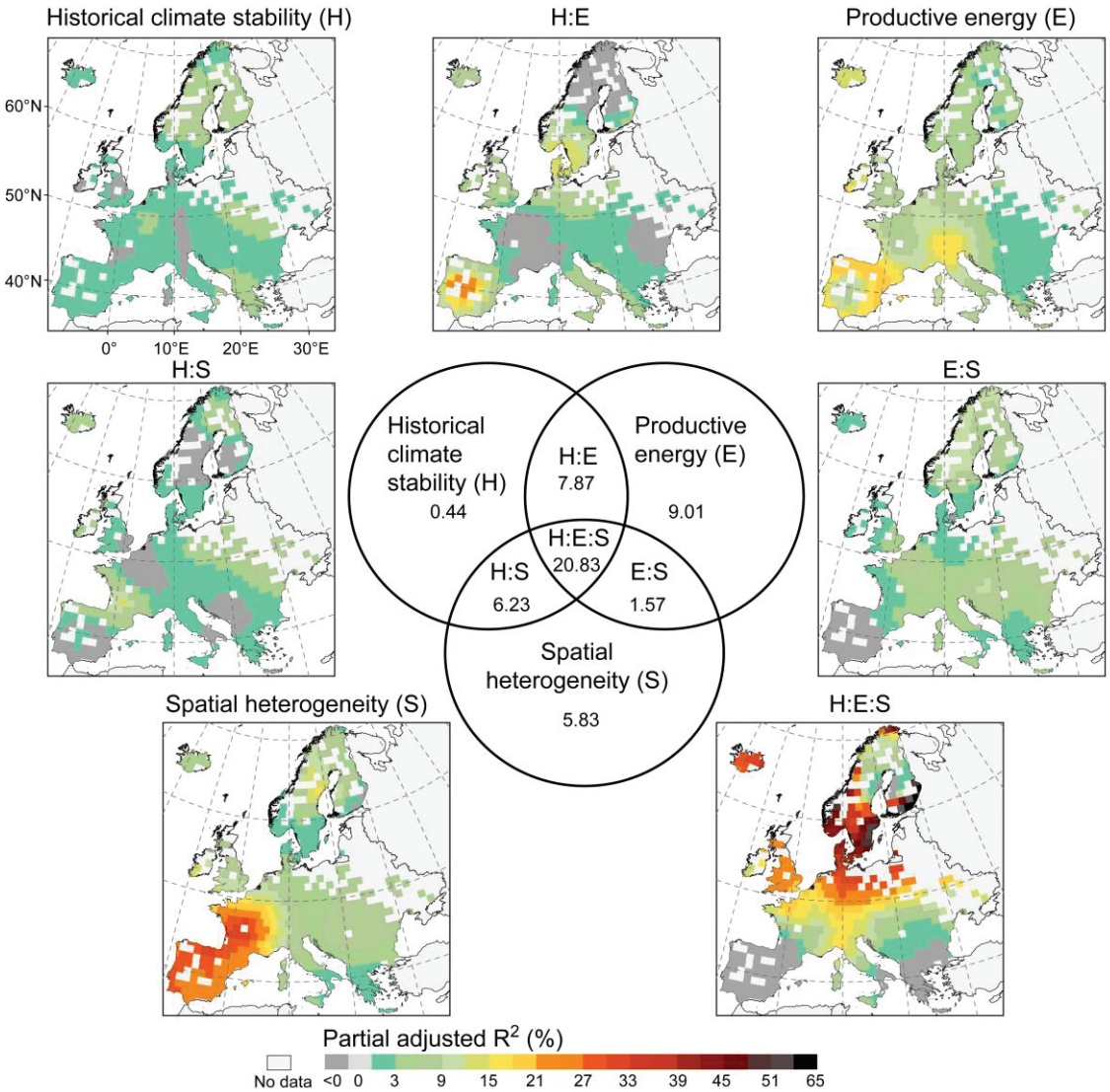


Figure 3 : Maps of proportions of variance explained by the independent and shared effects of historical climatic stability, productive energy and spatial energy, as provided by local partial adjusted R^2 of geographically weighted regression models. In abbreviations, colons denote shared variance between mechanisms. H:E: historical climatic stability and productive energy; H:S: historical climatic stability and spatial heterogeneity; E:S: productive energy and spatial heterogeneity; H:E:S= the three mechanisms. Venn diagram shows proportions of variance explained by the independent and shared effects of the three mechanisms, as provided by variance partitioning with ordinary least square models (see table 2).

Partial GWR models also pinpointed regions where patterns of species richness were not satisfactorily explained by any of the predictors considered. We propose several hypotheses which may account for the low explained variation in species richness in the Balkans and southern Italy: i) the failure to integrate other important predictors such as the complex paleogeography of these regions during the Tertiary -or biotic interactions (Schemske et al. 2009), ii) a high spatial heterogeneity in sampling effort among nearby cells, iii) a too coarse spatial resolution that is unable to sample steep ecological gradients in these regions (Davies et al. 2007).

Altogether, our findings emphasize the need for a broader perspective of groundwater biodiversity determinants at large spatial scale that explicitly considers the role of multiple mechanisms, their interactions, and changes in their relative contribution across space. As applied in this study, this broader perspective revives the importance of productive energy and spatial heterogeneity and their interaction in shaping crustacean diversity patterns in groundwater through their effect on diversification rates (i.e. the balance between speciation and extinction). It undermines the prominent role attributed to the disproportionate extinction of species in regions with high historical climate oscillations, because historical climate stability covaries with the two other mechanisms and productive energy in northern and southern Europe, respectively. Our inferences of groundwater biodiversity determinants are from correlation between species richness and distal predictors (e.g. AET). Yet, the importance of productive energy could further be tested by examining variation in diversification rates and DNA substitution rates among groundwater species-rich lineages that have colonized habitats with contrasted food supply (Evans et al. 2005, Fišer et al. 2008, Morvan et al. 2013). Similarly, identifying the ultimate causes of species richness patterns in regions where the effects of broad mechanisms covary would require multifaceted approaches to disentangle the role of niche-based mechanisms and dispersal constraints (Eme et al. 2014).

Acknowledgements

This research was funded by the European programme BioFresh (No. 226874), the French programme DEEP (No. ANR 08JCJC012001), and the PHC Proteus programme (No. 31199UM). CF and MZ were supported by Slovenian Research Agency (program P1-0184).

References

Araújo, M. B. et al. 2008. Quaternary climate changes explain diversity among reptiles and amphibians. – Ecography 31: 8–15.

Barton, K. 2013. MuMIn: Multi–model inference. R package version 1.9.0. Available at: http://CRAN.R-project.org/package=MuMIn

Beck, J. et al. 2012. What's on the horizon for macroecology? – Ecography 35: 673–683.

Bivand, R. 2012. spdep: Spatial dependence: weighting schemes, statistics and models. R package version 0.5-53. Available at: http://CRAN.R-project.org/package=spdep

Brown, J.H. 1995. Macroecology. – The University of Chicago Press.

Brown, J. H. et al. 2004. Toward a metabolic theory of ecology. – Ecology 85: 1771–1789.

Brunsdon, C. et al. 1996. Geographically Weighted Regression. – Geogr. Anal. 28: 281–298.

Burnham, K. P. and Anderson, D. R. 2002. Model selection and multimodel inference: a practical information–theoretic approach. – Springer.

Colwell, R. K. et al. 2004. The mid-domain effect and species richness patterns : what have we learned so far ? – Am. Nat. 163: E1–E23.

Cornu, J.-F. et al. 2013. The distribution of groundwater habitats in Europe. – Hydrogeol. J. 21: 949–960.

Culver, D. C. et al. 2006. The mid–latitude biodiversity ridge in terrestrial cave fauna. – Ecography 29: 120–128.

Currie, D. J. et al. 2004. Predictions and tests of climate–based hypotheses of broad–scale variation in taxonomic richness. – Ecol. Lett. 7: 1121–1134.

Datry, T. et al. 2005. Response of invertebrate assemblages to increased groundwater recharge rates in a phreatic aquifer. – J. North Am. Benthol. Soc. 24: 461–477.

Davies, R. G. et al. 2007. Topography, energy and the global distribution of bird species richness. – Proc. Biol. Sci. 274: 1189–1197.

Deharveng, L. et al. 2009. Groundwater biodiversity in Europe. – Freshw. Biol. 54: 709–726.

Dias, M. S. et al. 2013. Natural fragmentation in river networks as a driver of speciation for freshwater fishes. – Ecography 36: 683–689.

Dormann, C. F. et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. – Ecography 30: 609–628.

Dynesius, M. and Jansson, R. 2000. Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. – Proc. Natl. Acad. Sci. U. S. A. 97: 9115–9120.

Eiserhardt, W. L. et al. 2011. Testing the water–energy theory on American palms (Arecaceae) using geographically weighted regression. – PLoS One 6: e27027.

Eme, D. et al. 2013. Bayesian phylogeographic inferences reveal contrasting colonization dynamics among European groundwater isopods. – Mol. Ecol. 22: 5685–5699.

Eme, D. et al. 2014. Integrating phylogeography, physiology and habitat modelling to explore species range determinants. – J. Biogeogr. 41: 687–699.

Evans, K. L. et al. 2005. Species–energy relationships at the macroecological scale: a review of the mechanisms. – Biol. Rev. 80: 1–25.

Field, R. et al. 2009. Spatial species–richness gradients across scales: a meta–analysis. – J. Biogeogr. 36: 132–147.

Fišer, C. et al. 2008. A phylogenetic perspective on 160 years of troubled taxonomy of *Niphargus* (Crustacea: Amphipoda). – Zool. Scr. 37: 665–680.

Fišer, C. et al. 2012. Niche–based mechanisms operating within extreme habitats: a case study of subterranean amphipod communities. – Biol. Lett. 8: 578–81.

Foody, G. M. 2004. Spatial nonstationarity and scale–dependency in the relationship between species richness and environmental determinants for the sub–Saharan endemic avifauna. – Glob. Ecol. Biogeogr. 13: 315–320.

Foulquier, A. et al. 2010. Vertical change in dissolved organic carbon and oxygen at the water table region of an aquifer recharged with stormwater: biological uptake or mixing? – Biogeochemistry 99: 31–47.

Foulquier, A. et al. 2011. Surface water linkages regulate trophic interactions in a groundwater food web. – Ecosystems 14: 1339–1353.

Gibert, J. and Deharveng, L. 2002. Subterranean ecosystems: a truncated functional biodiversity. – Bioscience 52: 473–481.

Gouveia, S. F. et al. 2013. Nonstationary effects of productivity, seasonality, and historical climate changes on global amphibian diversity. – Ecography 36: 104–113.

Griebler, C. et al. 2014. Current developments in groundwater ecology – from biodiversity to ecosystem function and services. – Curr. Opin. Biotechnol. 27C: 159–167.

Hawkins, B. A. et al. 2003. Energy, water, and broad–scale geographic patterns of species richness. – Ecology 84: 3105–3117.

Hijmans, R. J. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. – Int. J. Climatol. 25: 1965–1978.

Hof, C. et al. 2008. Latitudinal variation of diversity in European freshwater animals is not concordant across habitat types. – Glob. Ecol. Biogeogr. 17: 539–546.

Hortal, J. et al. 2013. Species richness can decrease with altitude but not with habitat diversity. – Proc. Natl. Acad. Sci. U. S. A. 110: E2149–E2150.

Kerr, J. T. and Packer, L. 1997. Habitat heterogeneity as a determinant of mammal species richness in high–energy regions. – Nature 385: 252–254.

Kerr, J. T. et al. 2001. Remotely sensed habitat diversity predicts butterfly species richness and community similarity in Canada. – Proc. Natl. Acad. Sci. U. S. A. 98: 11365–11370.

Kissling, W. D. and Carl, G. 2007. Spatial autocorrelation and the selection of simultaneous autoregressive models. – Glob. Ecol. Biogeogr. 17: 59–71.

Legendre, P. and Legendre, L. 1998. Numerical Ecology. – Elsevier.

Leprieur, F. et al. 2011. Partitioning global patterns of freshwater fish beta diversity reveals contrasting signatures of past climate changes. – Ecol. Lett. 14: 325–334.

Morueta–Holme, N. et al. 2013. Habitat area and climate stability determine geographical variation in plant species range sizes. – Ecol. Lett. 16: 1446–1454.

Morvan, C. L. M. et al. 2013. Timetree of Aselloidea reveals species diversification dynamics in groundwater. – Syst. Biol. 62: 512_522.

O'Brien, E. M. 2006. Biological relativity to water-energy dynamics. – J. Biogeogr. 33: 1868–1888.

Oberdorff, T. et al. 2011. Global and regional patterns in riverine fish species richness: A Review. – Int. J. Ecol. 2011: 1–12.

Ohlemüller, R. et al. 2008. The coincidence of climatic and species rarity: high risk to small–range species from climate change. – Biol. Lett. 4: 568–572.

Oksanen, J. et al. (2012). vegan: Community Ecology Package. R package version 2.0-5. Avalaible at: http://CRAN.R-project.org/package=vegan

Palmer, M. W. 1994. Variation in species richness: towards a unification of hypotheses. – Folia Geobot. Phytotaxon. 29: 511–530.

Qian, H. and Ricklefs, R. E. 2000. Large-scale processes and the Asian bias in species diversity of temperate plants. – Nature 407: 180–182.

Rahbek, C. and Graves, G. R. 2001. Multiscale assessment of patterns of avian species richness. – Proc. Natl. Acad. Sci. U. S. A. 98: 4534–4539.

Rangel, T. F. et al. 2010. SAM: a comprehensive application for spatial analysis in macroecology. – Ecography 33: 46–50.

R Development Core Team (2013) R: A language and environment for statistical computing. – R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org

Ruggiero, A. and Kitzberger, T. 2004. Environmental correlates of mammal species richness in South America: effects of spatial structure, taxonomy and geographic range. – Ecography 27: 401–416.

Sastre, P. et al. 2009. A Geoplatform for improving accessibility to environmental cartography. – J. Biogeogr. 36: 568–568.

Schemske, D. W. et al. 2009. Is there a latitudinal gradient in the importance of biotic interactions? – Annu. Rev. Ecol. Evol. Syst. 40: 245–269.

Stoch, F. and Galassi, D. M. P. 2010. Stygobiotic crustacean species richness: a question of numbers, a matter of scale. – Hydrobiologia 653: 217–234.

Strayer, D. L. et al. 1997. Oxygen, organic matter, and sediment granulometry as controls on hyporheic animal communities. – Archiv Hydrobiol. 140:131–44.

Svenning, J.–C. et al. 2009. Plio–Pleistocene climate change and geographic heterogeneity in plant diversity–environment relationships. – Ecography 32: 13–21.

Tisseuil, C. et al. 2013. Global diversity patterns and cross–taxa convergence in freshwater systems. – J. Anim. Ecol. 82: 365–376.

Trabucco, A. and Zomer, R.J. 2010. Global Soil Water Balance Geospatial Database. CGIAR Consortium for Spatial Information. Published online, available from the CGIAR–CSI GeoPortal at: http://www.cgiar-csi.org/data/global-high-resolution-soil-water-balance

Trontelj, P. et al. 2009. A molecular test for cryptic diversity in ground water: how large are the ranges of macro–stygobionts? – Freshw. Biol. 54: 727–744.

Trontelj, P. et al. 2012. Ecomorphological convergence of cave communities. – Evolution 66: 3852–3865.

Urban, M. C. et al. 2008. The evolutionary ecology of metacommunities. – Trends Ecol. Evol. 23: 311–317.

Voght et al. 2007. A pan–European river and catchment Database. European Commission, Joint Research Centre, DOI 10.2788/35907. Available at: http://ccm.jrc.ec.europa.eu/php/index.php?action=view&id=23

Wheeler, D. and Tiefelsdorf, M. 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. – J. Geogr. Syst. 7: 161–187.

Whittaker, R. J. et al. 2001. Scale and species richness: towards a general, hierarchical theory of species diversity. – J. Biogeogr. 28: 453–470.

Wiens, J. J. 2011. The causes of species richness patterns across space, time, and clades and the role of "ecological limits." – Q. Rev. Biol. 86: 75–96.

Willig, M. R. et al. 2003. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. – Annu. Rev. Ecol. Evol. Syst. 34: 273–309.

Zagmajster, M. et al. 2014. Geographic variation in range size and beta diversity of groundwater crustaceans: insights from habitats with low thermal seasonality. – Submitted to Global Ecol. Biogeogr.

Zuur, A. et al. 2010. Mixed Effects Models and Extension in Ecology with R. – Springer.

Chapitre 3) Emphase sur le processus de dispersion par l'étude de la dynamique des aires de répartition à travers une approche de phylogéographie comparative

Les articles 2 et 3 du chapitre précédent mettent en évidence une forte décroissance de la richesse spécifique au nord de l'Europe et soutiennent la règle de Rapoport, c'est-à-dire une augmentation de la taille moyenne des aires de répartition avec la latitude. Ces résultats suggèrent qu'un faible nombre d'espèces aurait eu les capacités de disperser afin de coloniser le nord de l'Europe. Cette colonisation serait intervenue récemment lors de périodes de retrait glaciaire au Pléistocène. Cette suggestion est également confortée par les patrons de béta diversité qui indiquent une augmentation latitudinale de la contribution de la diversité emboitée à la diversité béta totale. En revanche, les corrélogrammes de Mantel montrent que les communautés sont presque entièrement renouvelées sur des distances inférieures à 500 km. Ce résultat suggère que la dispersion qui aurait permis une extension des aires de répartition vers le nord serait intervenue à partir de refuges situés au centre de l'Europe plutôt qu'à partir des péninsules de l'Europe du sud (Hewitt, 1996).

Ce chapitre sous la forme d'un article a pour objectif d'évaluer plus précisément le rôle du processus de dispersion et notamment celui des colonisations postglaciaires sur les patrons d'aire de répartition. Pour ce faire, il utilise des outils moléculaires et une approche de phylogéographie comparative menée sur cinq espèces morphologiques d'isopodes aquatiques souterrains présentant de larges aires de répartition situées dans des régions impactées par les glaciers et/ou les climats froids du Pléistocène.

3.1) Article 4 : Bayesian phylogeographic inferences reveal contrasting colonization dynamics among European groundwater isopods

D. Eme[1], F. Malard[1], L. Konecny-Dupré[1], T. Lefébure[1] and C. J. Douady[1,2] *

[1]: Université de Lyon ; UMR5023 Ecologie des hydrosystèmes Naturels et Anthropisés ; Université Lyon1 ; ENTPE ; CNRS ; 6 rue Raphaël Dubois, 69622 Villeurbanne, France.
[2]: Institut Universitaire de France, Paris, F-75005, France.

Article publié dans *Molecular Ecology* **22**, 5685–5699.

Abstract

The potentially important role of northern microrefugia during post-glacial dispersal is challenging the view of Southern Europe as a refuge and source area of European biota. In groundwaters, large geographic ranges of presumably good dispersers are increasingly suspected to consist of assemblages of cryptic species with narrow ranges. Moreover, a large species range, even when confirmed by molecular evidence, tells us little about the spatiotemporal dynamics of dispersal. Here, we used phylogenetic inferences, species delineation methods and Bayesian phylogeographic diffusion models to test for the likelihood of post-glacial colonization from distant refugia among five morphospecies of *Proasellus* (Isopoda, Asellidae). All morphospecies except one were monophyletic, but they comprised a total of 15 - 17 cryptic species. Three cryptic species retained ranges that spanned a distance > 650 km, similar to that of the nominal morphospecies. Bayesian diffusion models based on mitochondrial markers revealed considerable spatiotemporal heterogeneity in dispersal rates, suggesting that short-time dispersal windows were instrumental in shaping species ranges. Only one species was found to experience a recent, presumably post-glacial, range expansion. The Jura and Alpine foothills probably played a major role in maintaining diversity within *Proasellus* in northern regions by acting both as diversification hotspots and Pleistocene refugia. Gaining insight into the spatiotemporal heterogeneity of dispersal rates revealed contrasting colonization dynamics among species that were not consistent with a global postglacial colonization of Europe from distant refugia.

Keywords species range dynamics, dispersal rates, cryptic diversity, Pleistocene refugia, groundwater crustaceans, Europe.

Introduction

Species distributions at continental scales are ultimately caused by the interactions among dispersal, speciation and extinction (Ricklefs 2004; Mittelbach *et al.* 2007). In regions of constantly changing climate, dispersal is of primary importance because the ability of vagile and generalist species to colonize vacant habitats should theoretically be selected (Dynesius & Jansson 2000). Earlier phylogeographic studies in Europe effectively supported paleobiogeographic scenarios that implied recent dispersal events from distant refugia (Hewitt 1996, 2000; Taberlet *et al.* 1998). During interglacials, Alpine/Arctic species experienced severe range contraction into high-elevation or northern refugia, whereas the ranges of Mediterranean and continental species expanded from refugia located in southern and eastern Europe, respectively (review in Schmitt 2007). However, subsequent phylogeographic studies, some using environmental niche modeling, documented recent dispersal events from multiple *in situ* northern microrefugia; i.e. small areas in which environmental conditions differed favorably from the surrounding environment (Stewart & Lister 2001; Provan & Bennett 2008; Cordelier & Pfenninger, 2010). The potentially important role of northern microrefugia during post-glacial colonization is challenging the generally held view of Southern Europe as a long-term refuge and source area for European biota, even among the best-studied organisms such as mammals (Svenning *et al.* 2011).

Distribution patterns of obligate groundwater species in Europe undoubtedly retain the imprint of cold Pleistocene climate because regional species richness dramatically declines northward in formerly glaciated and permafrost areas (Hof *et al.* 2008; Stoch & Galassi 2010). The comparatively wider ranges of groundwater species in northern Europe suggest post-glacial colonization from distant refugia in non-permafrost areas (Henry 1976; Malard *et al.* 2009; Stoch & Galassi 2010). Yet, whether wide ranges are evidence of dispersal is being severely questioned by the discovery of highly divergent evolutionary units – hereafter referred to as cryptic species – within widely-distributed groundwater morphospecies (Lefebure *et al.* 2006a, 2007; Finston *et al.* 2007; Zakšek *et al.* 2009; Morvan *et al.* 2013). Trontelj *et al.* (2009) reviewed cryptic lineage diversity within 14 widely-distributed groundwater species and suggested that ranges with extents higher than 200 km were extremely rare in groundwaters. If so, this would suggest that local dispersal from *in situ* northern microrefugia, including local ice-free mountain tops, (Lefébure *et al.* 2007), unfrozen groundwater beneath the ice (Holsinger 1980) and periglacial habitats (Foulquier *et al.* 2008) have played a major role during post-glacial colonization. However, the discovery of cryptic species within widely-distributed morphospecies alone is insufficient to reject dispersal from distant refugia, unless exhaustive spatial sampling shows that cryptic species have a much narrower distribution range than the nominal morphospecies. Lefébure *et al.* (2006a) revealed three highly divergent cryptic species within the groundwater amphipod *Niphargus virei* Chevreux, 1896, but the range of one of these still spanned a distance of 700 km, similar to that of the nominal morphospecies.

Geographic range size tells us little about the dynamics of dispersal. Although range expansion in groundwaters proceeds by stepping stone dispersal, dispersal can be extremely heterogeneous over time and space if it preferentially occurs during short-time environmental windows of increased habitat connectivity such as periods of intense post-glacial sediment deposition (Ward & Palmer, 1994). Evidence in favor of recent range expansion can preferably come from Bayesian phylogeographic diffusion models that infer the spatiotemporal heterogeneity of dispersal rates over the course of clade evolution (Lemey *et al.* 2010; Pybus *et al.* 2012). These models simultaneously reconstruct the evolutionary history and ancestral geographic locations of a clade in a

continuous landscape using either a homogenous diffusion process (random walk) or a heterogeneous diffusion process (relaxed random walk), which enables bursts of dispersal.

In this study, we tested for the likelihood of post-glacial colonization from distant refugia among five widely-distributed morphospecies of *Proasellus* (Isopoda, Asellidae) using two mitochondrial genes (*COI* and *16S*) and one nuclear gene (*28S*). First, we predicted that some cryptic species would retain large geographic ranges if Pleistocene climatic oscillations selected for vagility (see Dynesius & Jansson 2000). This prediction was tested by delineating cryptic species using phylogenetic inferences and molecular species delineation methods. Second, we predicted that dispersal rates would be extremely heterogeneous over the course of a species' evolution because dispersal might have preferentially occurred during short-time windows of increased habitat connectivity. Third, the increase in dispersal rates during short-time windows might have been sufficiently strong to enable post-glacial colonization of large areas from distant refugia. To test for the second and third predictions, we assessed the rate and timing of dispersal and location of refugia using Bayesian phylogeographic spatial diffusion models.

Materials and methods

Species selection and sampling

The genus *Proasellus* (Pancrustacea, Isopoda) is one of the most diverse and most widely-distributed genera of groundwater organisms in Europe. In this genus, five morphospecies, hereafter referred to as focal species, were selected according to the two following criteria (Fig. 1). First, they had wide distribution ranges with maximum linear extent > 200 km (MLE: the straight-line distance between the two most distant known localities). Second, part of their present-day distribution range extended into areas that were covered by ice or continuous permafrost during the Last Glacial Maximum. Of these five morphospecies, *Proasellus cavaticus* (Leydig, 1871), *Proasellus strouhali* (Karaman, 1955), *Proasellus synaselloides* (Henry, 1963) and *Proasellus walteri* (Chappuis, 1948) belong to the Alpine lineage that contains 34 obligate groundwater species and a single surface water species (Morvan *et al.* 2013). *Proasellus slavus* (Remy, 1948) belongs to the Slavus lineage, which consists exclusively of obligate groundwater species. The geographic ranges of species were delineated with nearest neighbor convex hull method using morphospecies occurrences from the authors' data base. We selected the minimum number of neighbors that provided a continuous range. For each focal morphospecies, sampling was optimized to maximize the probability of detecting cryptic diversity (Table S1). Whenever possible, samples were obtained from localities: 1) located at cardinal points of morphospecies range; 2) belonging to distinct river catchments (i.e. sea outlets); 3) harboring isolated populations; and 4) containing populations previously described as morphological subspecies (see Table S1). Number of sampling sites ranged from 8 to 25 among species (Fig. 1). Samples were collected from caves, springs, wells, and the hyporheic zone of streams. They were placed in 96% ethanol at ambient temperature for transport back to the laboratory, then at 4°C until sorting and morphological identification. Individuals were identified using original species diagnoses, which were mostly based on the morphology of male copulatory organs (second pleopod). Male pleopods were mounted on slides for identification purposes and the remaining part of specimens was conserved at -20°C until molecular analysis.

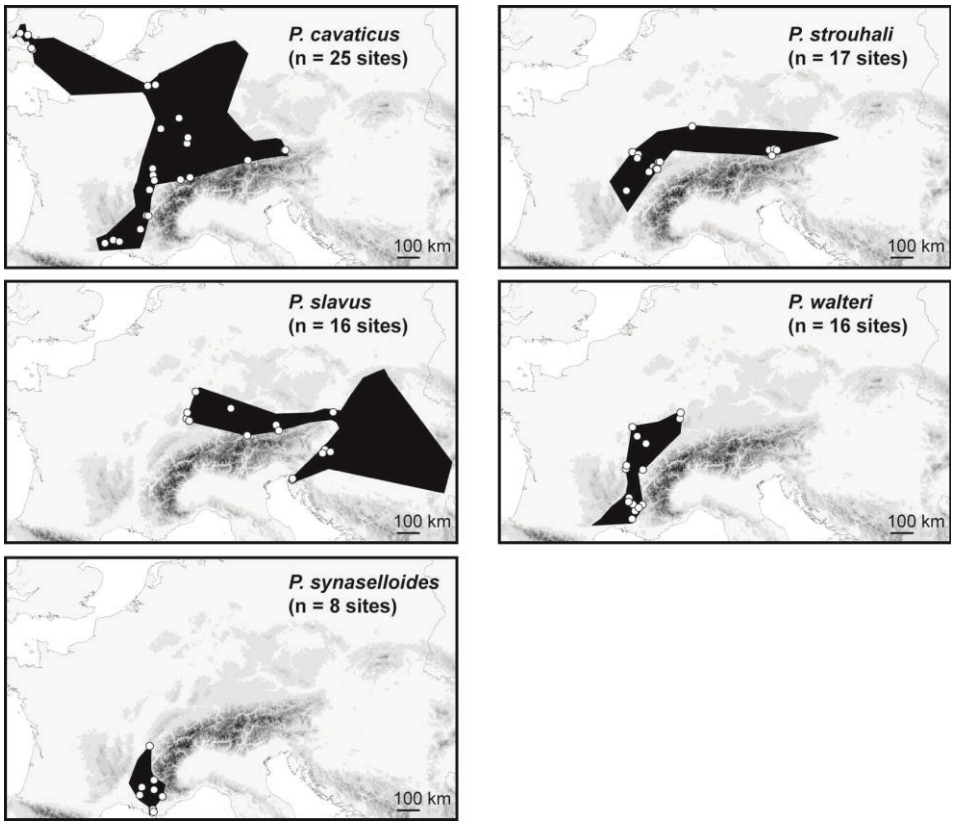


Figure 1: Geographic ranges of the five *Proasellus* morphospsecies showing the location of sampling sites (white dots).

Molecular data acquisition

Molecular data were obtained following protocols described in Calvignac *et al.* (2011) and Morvan *et al.* (2013) (Table S2). Briefly, genomic DNA was extracted from 3 specimens, whenever possible, for each locality using a chloroform DNA extraction protocol. Two mitochondrial (*COI*, *16S*) and one nuclear (*28S*) gene fragments were obtained using for each a combination of different primer pairs (*COI*, *16S* and *28S*, see Table S2), long range PCR (*COI* and *16S*) and pre-PCR dilution of genomic DNA (*COI* and *16S*) to prevent misleading inclusion of nuclear mitochondrial pseudogenes (*COI* and *16S*) or paralogs (*28S*). Sanger sequencing was performed by service providers (GATC Biotech; Konstanz, Germany; Eurofins MWG Operon, Ebersberg, Germany; BIOFIDAL; Vaulx-en-Velin, France). Chromatograms were visualized using FINCHTV version 1.4.0 (Geospiza, Inc.; Seattle, WA, USA; http://www.geospiza.com). Sequences were submitted to GenBank (accession numbers KC6100479-KC610500; KC610160-KC610269; KC610369-KC610418).

Data analysis

Monophyly analysis

Monophyly of the 5 focal morphospecies was assessed using large-scale phylogenetic inferences under likelihood and Bayesian frameworks. Our dataset comprised a total of 361 DNA sequences from 149 individuals (Table 1). Within each morphospecies, individuals with identical sequences for the 3 loci were collapsed using a custom made Perl script (Morvan *et al.* 2013) and a single best representative was retained in subsequent analyses. The best representative was defined as

the longest and less ambiguous concatenated sequence using R software (R Development Core Team 2011) and seqinr package (version 3.0-3; Charif & Lobry 2007). Some 187 sequences (*COI*, *16S*, *28S*) belonging to 63 individuals representing 59 non-focal morphospecies of *Proasellus* and 4 outgroup species were included to challenge the monophyly of focal morphospecies (Table S1). The COI, 16S and 28S genes were individually aligned with MUSCLE (Edgar 2004) as implemented in SEAVIEW (version 4.2.12; Gouy *et al.* 2010). Poorly aligned positions and divergent regions were removed using GBLOCKS (version 0.91b; Castresana 2000). Alignments were declared in GBLOCKS as DNA (*16S* and *28S*) or codon (*COI*) and parameters were set following author's recommendation for less stringent selection. Selections were then checked by eye in SEAVIEW.

We tested for congruence between the three gene trees before using a single partition in PHYML (version 3.0; Guindon *et al.* 2010) or a linked topology among 3 independent partitions in BEAST (version 1.6.2; Drummond & Rambaut 2007). Non-parametric bootstrap analysis was used to search for strongly supported conflicting bipartitions between any two gene trees (Lefébure & Stanhope 2007). Support for each bipartition was obtained by bootstrapping a maximum likelihood (ML) tree search using PHYML (GTR+G+I model of evolution, 200 pseudo replicates). A bipartition was considered supported if it had a bootstrap support ≥90%. This analysis was performed on a subset of 134 individuals representing 66 morphospecies for which the three loci were available.

Most likely topology was inferred using PHYML under the AIC best-fitted model of evolution (GTR+G+I) selected by JMODELTEST (Posada 2008), a BIONJ as starting tree and the best of NNI and SPR as tree search heuristics. In the maximum likelihood analysis, the three loci (*COI*, *16S*, *28S*) were treated both individually and as a single partition. Parameters of the evolution model were estimated directly with PHYML by maximum likelihood. Topology robustness was assessed using 500 and 1000 bootstrap pseudo-replicates for the individual loci and single partition, respectively. We considered strong node support when bootstrap support (bs) was superior to 90%.

The most probable chronogram was inferred with BEAST using the AIC best-fitted model of evolution (GTR+G+I) selected individually for each locus by JMODELTEST. We performed three independent runs for each locus (*COI*, *16S*, *28S*) and one run in which the three loci and positions (*COI*) were treated as independent partitions with a linked topology. For all runs, a Yule model was selected as the speciation model because we had no access to extinction rates (absence of fossils). We used an uncorrelated Lognormal Bayesian relaxed molecular clock because the posterior probability of the ucld.stdev parameter in BEAST did not abut against zero (see recommendations in Drummond *et al.* 2007). Paleogeographic calibration points between outgroup species were used to constrain the age of two nodes. The definition of priors for maximum clade credibility (MCC) tree inference, including that of calibration points, is further detailed in Table S3. The Markov Chain Monte Carlo (MCMC) was run for 100 million generations and sampled every 10,000 generations. Three independent runs were first checked for convergence with TRACER (version 1.5; Rambaut & Drummond 2009) and then combined with LOGCOMBINER, discarding the first 10 million iterations of each run as burn-in. The posterior sample of the trees obtained was summarized by TREEANNOTATOR to produce a MCC tree and displayed with FIGTREE (version 1.3.1 http://tree.bio.ed.ac.uk/software/figtree/). Topology robustness was assessed using posterior probability, strong node support being considered when posterior probability (pp) was superior or equal to 0.99.

Molecular species delineation

Cryptic species were delineated based on the COI gene using the *COI* threshold developed by Lefébure *et al.* (2006b) and the general mixed Yule-coalescent (GMYC) model of Pons *et al.* (2006). In addition, we examined the distribution of 28S nuclear gene haplotypes among cryptic species. The *COI* threshold is based on the observation made from 1500 *COI* sequences of 276 crustacean morphospecies that two monophyletic groups diverging by more than 0.16 substitution per site, as measured by patristic distances, have a strong probability of belonging to different species (Lefébure *et al.* 2006b). The GMYC model delineates species from branching rates in mixed population-phylogenetic trees without any prior definition of populations or species.

All focal *COI* sequences (n=149) and 63 sequences corresponding to non-focal *Proasellus* and outgroup species were aligned, cleaned and collapsed, as described above in the monophyly section. For the threshold method, most likely topology was inferred using PHYML. Patristic distances were extracted using R package ape (version 2.7-3; Paradis *et al.* 2004) and cryptic species were delimited using R package cluster (version 2.7-3; Maechler *et al.* 2012). For the GMYC method, MCC tree was inferred using BEAST (cf. monophyly section for settings). The MCMC was run for 200 million generations and sampled every 20,000 generations. Five independent runs were first checked for convergence with TRACER and the best chain with the highest effective sample size (ESS) was retained after discarding the first 100 million iterations as burn-in. The posterior sample of the trees obtained was summarized by TREEANNOTATOR to produce a MCC tree. Finally, species were defined from this tree using the single-threshold GMYC method (Pons *et al.* 2006) as implemented in R package splits (version 1.0-11; Ezard *et al.* 2009).

Bayesian spatial diffusion analysis

Bayesian phylogeographic diffusion models were used to infer phylogeographic history in continuous space while accounting for phylogenetic uncertainty (Lemey *et al.* 2010). At this stage of the analysis, we only considered three species delineated by molecular methods that retained wide distribution ranges (MLE > 200 km) extending into formerly glaciated or permafrost areas. Georeferenced DNA sequences were used to reconstruct genealogy between individuals and the geographic location (spatial coordinates) of ancestors along the tree. To infer ancestor locations, we performed both random and relaxed random walks, which respectively accommodated constant and relaxed variance of dispersal rates among branches of the tree.

Molecular data corresponding to a total of 193 sequences were assembled and analyzed separately for each species using the *COI*, *16S* and *28S* loci (for details see species referred to as *P.* ca1, *P.* st1, *P.* sl1 in Table S1). However, our phylogeographic inferences were mainly based on mitochondrial genes because the 28S gene was poorly informative at the intraspecific level. Individuals with identical haplotypes were retained as their associated coordinates conveyed geographical information. Alignments were performed as previously described. Bayesian spatial diffusion analyses were performed under BEAST. The three loci were treated as independent partitions but with linked topology. We used GTR+G+I models as substitution models, non-parametric Bayesian skyride plots (Minin *et al.* 2008) as coalescent tree models, strict molecular clock models (ucld.stdev parameter abutting against zero) and three spatial diffusion models. Priors for the ages of the most recent common ancestors (MRCA) of the three species were defined according to previous age estimates (i.e. BEAST chronogram inferred for testing monophyly).

As diffusion models (or dispersal kernel, see Nathan *et al.* 2012), we considered a random walk model following a Brownian motion process and two relaxed random walk models using a

Gamma and Lognormal dispersal kernel (Lemey *et al.* 2010). Uncertainty in the spatial diffusion models was taken into account by bivariate precision matrices. Jitter option was enforced to add random noise to identical coordinates because duplicate haplotypes could occur at neighboring localities (Table S3). For each spatial diffusion model and species, 6 independent MCMCs were run during 200 million generations and sampled every 20,000 generations. Convergence was evaluated with TRACER and the best chain with the highest ESS for each spatial diffusion model was retained after discarding the first 10 million iterations as burn-in. The best spatial diffusion models were selected using Bayes factors (BF) (Kass & Raftery 1995). BFs were estimated with TRACER (version 1.5; Rambaut & Drummond 2009) using the harmonic mean estimator of the marginal likelihood of the models (Suchard *et al.* 2001). Interpretation of Bayes factors was done according to Kass & Raftery (1995). Namely, support of model 1 against model 0 was considered positive when 2 ln (BF {1-0}) >2, strong for values >6 and very strong for values >10. The posterior samples of the trees obtained were summarized by TREEANNOTATOR to produce MCC trees. Dispersal rates and historical range dynamics of the three species were subsequently inferred by extracting spatial coordinates and ages of nodes from the best diffusion model outputs in BEAST. The dispersal rate *D* of branch *i* was determined as follow:

$$D_i = d_i/h_i$$

where d_i is the great circle distance between the two ends of branch i, and h_i represents the time corresponding to the length of branch i. We used relative time units (RTU) by expressing time as a percentage of the age of the oldest MRCA of the three species. The MRCA corresponds to the diversification node of a clade as estimated from the BEAST chronogram used to test monophyly. This test allowed comparing dispersal rates between species without making strong assumptions on absolute age estimation. Geographic distances were computed using rdist.earth command from fields v6.6.3 R package (Furrer *et al.* 2012). Species range dynamics were displayed by projecting MCC tree branches on a map and color gradients were used to represent the relative age of expansion phases.

Results

Monophyly of focal morphospecies

The analysis of phylogenetic congruence among gene trees revealed no conflicting bipartition at a bootstrap support level of 90%. All focal morphospecies except *P. slavus* were found to be monophyletic both by the most likely topology using a single partition and the MCC tree with linked topology (Table 1, Fig. 2). This result was corroborated by independent gene trees built either with PHYML or BEAST. The 28S gene did not recover the monophyly of *P. synaselloides* but branch supports were very low (Table S4, Fig. S1). *P. slavus* was a paraphyletic group made of four highly divergent lineages (*P.* sl1-4; Fig. 2). Lineage *P.* sl1 and *P. nolli* formed a strongly supported monophyletic group (bs = 100%; pp = 1) with a weak genetic divergence between populations (mean pairwise patristic distances of 0.03 and 0.01 substitution/site for the *COI* and *16S*, respectively). Despite being monophyletic, the four other morphospecies showed a high within species genetic divergence suggesting cryptic diversity (see genetic divergences in Table S4 and Fig. S2).

Table 1 Molecular data set for the five focal morphospecies and results of monophyly analysis and molecular species delineation. bs is boostrap support for the ML tree with a single partition and pp is

posterior probability for the MCC tree with a linked topology. Parap. is for paraphyletic morphospecies.

Morphospecies name	Number of individuals	Number of sequences			Monophyly support		Number of cryptic species	
		COI	*16S*	*28S*	bs (%)	pp	*COI* threshold	GMYC
P. cavaticus	43	43	35	26	100	1	4	2
P. strouhali	24	24	23	14	100	1	2	2
P. slavus	32	32	31	16	Parap.	Parap.	4	4
P. synaselloides	18	18	18	6	96.9	1	4	4
P. walteri	32	32	30	13	100	1	3	3
Total	149	149	137	75			17	15

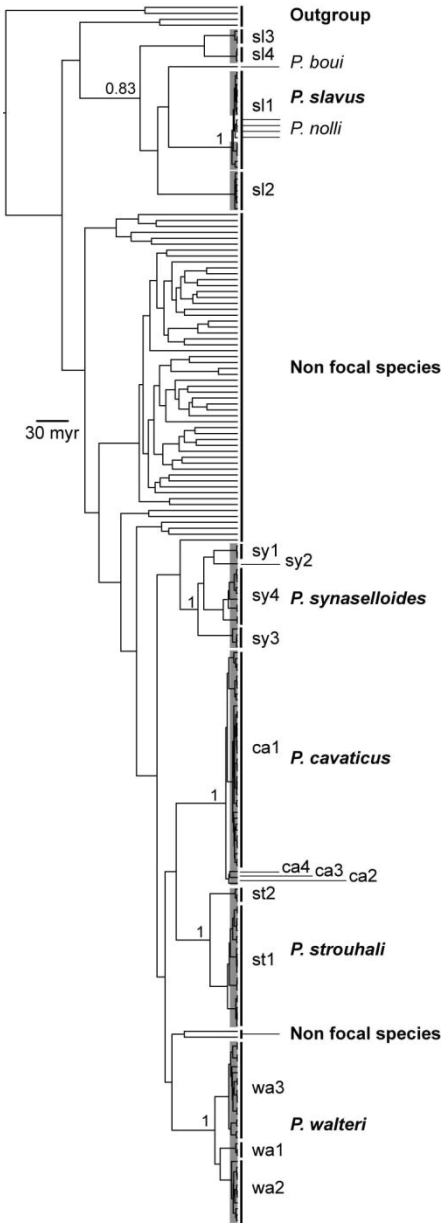


Figure 2: Maximum clade credibility tree of 68 Aselloidea morphospecies used to test for monophyly of the five morphospecies (grey patterns): *P. cavaticus*, *P. strouhali*, *P. slavus*, *P. synaselloides* and *P. walteri*, (BEAST analysis based on COI, 16S and 28S genes with a linked topology and an uncorrelated lognormal relaxed molecular clock model). Each terminal branch for non-focal and outgroup species corresponds to a morphospecies. Abbreviations ca, st, sl, sy and wa indicate cryptic species delineated using the COI gene. Numbers along selected branches are posterior probabilities.

Crytpic diversity within focal morphospecies

The threshold and GMYC methods recognized 17 and 15 cryptic species, respectively, thereby tripling species richness (Table 1). Haplotype clustering was identical between the two species delineation methods, with the exception of *P. cavaticus*. In the latter morphospecies, the GMYC method assigned three haplotypes (i.e. hb_17, hb_18, hb_23; see Table S1) to a single cryptic species, whereas these haplotypes were attributed to three distinct cryptic species by the threshold method (*P.* ca2-4). Species delineation based on mitochondrial DNA was partly corroborated by the distribution of *28S* haplotypes among cryptic species. Nine cryptic species out of 16 identified by the *COI* threshold method and 10 cryptic species out of 14 identified by the GMYC method had their own sets of *28S* haplotypes (Fig. S3). One cryptic species, *P. synaselloides* (sy2), could not be evaluated.

Morphospecies showing the smallest distribution ranges (i.e. *P. synaselloides* and *P. walteri*) contained as many or even more cryptic species than the most-widely distributed species. Overall,

cryptic species did not recover morphological subspecies (see Table S1). Both species delineation methods aggregated all haplotypes of *P. nolli* into a single cryptic species of *P. slavus* (*P.* sl1), confirming the inappropriate morphological distinction of *P. nolli* revealed earlier by phylogenetic inferences (see above).

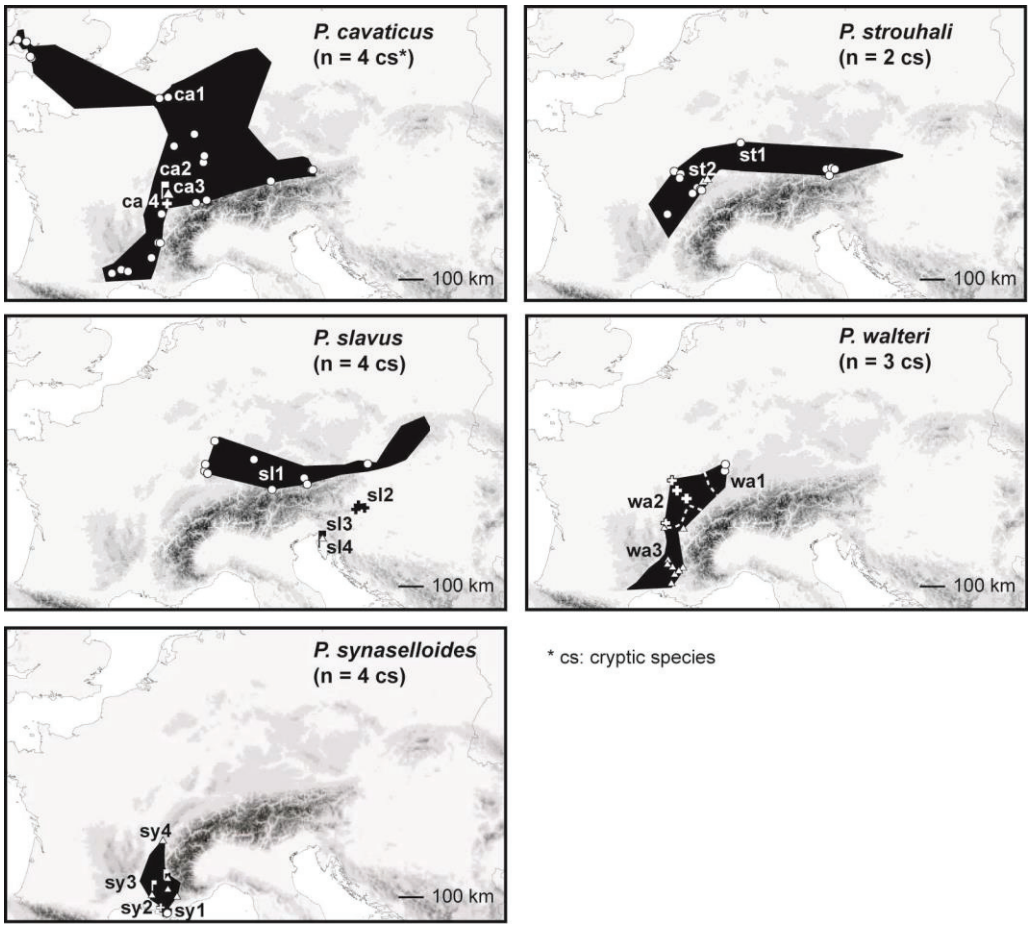


Figure 3: Distribution of cryptic species. Abbreviations for cryptic species are the same as in Table 2 and figure 2. Black polygon in *P. cavaticus*, *P. strouhali* and *P. slavus* show the geographic ranges of the most widely distributed cryptic species (white dots). Geographic ranges of cryptic species are indicated with dashed lines for *P. walteri*; they are not shown for *P. synaselloides*.

Cryptic species within the three most widely-distributed morphospecies were not homogeneously distributed across space (Fig. 3). Each morphospecies comprised one widely-distributed cryptic species (i.e. *P.* ca1, *P.* st1, and *P.* sl1) as well as a set of narrowly-distributed cryptic species located at the northern (i.e. *P.* ca2-4 and *P.* st2 in the Jura) and southern periphery of the Alps (i.e. *P.* sl2-4 in the foothills of Slovenian Alps). Consequently, the MLE of the most widely-distributed cryptic species was similar to that of the morphospecies (Table 2). In contrast, all cryptic species within *P. walteri* and *P. synaselloides* had a much smaller geographic range than the nominal morphospecies. Overall, cryptic diversity analysis suggested a pattern of increasing range size with increasing latitude. Widely- and narrowly-distributed cryptic species distinctly occurred at northern and southern latitudes, respectively (Table 2, Fig. 3).

Table 2 Linear extent and centroid of geographic ranges of nominal morphospecies and cryptic species (*P.* ca, *P.* st, *P.* sl, *P.* sy and *P.* wa). Abbreviations for cryptic species are the same as in figures 2 and 3. s.s. is for a single site species.

Species name	Maximum linear extent (km)	Species centroid	
		Latitude*	Longitude*
P. cavaticus	1312	47.82	4.94
P. ca1	1312	47.82	4.94
P. ca2	s.s.	47.01	5.59
P. ca3	s.s.	46.74	5.64
P. ca4	s.s.	46.53	5.73
P. strouhali	704	47.43	9.02
P. st1	704	47.43	9.02
P. st2	6	47.32	6.40
P. slavus	667	47.37	12.32
P. sl1	663	48.31	12.06
P. sl2	36	46.63	15.93
P. sl3	s.s.	45.60	13.85
P. sl4	s.s.	45.54	13.83
P. synaselloides	300	44.60	5.77
P. sy1	s.s.	43.26	5.96
P. sy2	s.s.	43.37	5.92
P. sy3	64	44.40	5.56
P. sy4	235	44.93	5.77
P. walteri	530	46.12	6.09
P. wa1	27	48.21	7.62
P. wa2	192	46.79	5.09
P. wa3	229	44.92	5.22

*Expressed in decimal degrees (N and E).

Heterogeneity of dispersal rates

Bayesian phylogeographic inferences were used to reconstruct the range dynamics of the three widely distributed cryptic species, namely *P.* ca1 (with *P.* ca2-4 as outgroup), *P.* st1 (with *P.* st2 as outgroup), and *P.* sl1 (without outgroup due to uncertainty and high divergence of sister species). For the three species, selection of spatial diffusion models using Bayes factors strongly supported relaxed random walk models inferring heterogeneous dispersal rates among branches against the Brownian random walk model inferring homogeneous dispersal rates (table 3, see also Fig. S4). Among the two relaxed random walked models, support for the Gamma kernel against the lognormal dispersal kernel was very strong for *P.* ca1 (2ln BF = 104.02) and *P.* sl1 (2ln BF = 36.54), and positive for *P.* st1 (2ln BF = 4.04) (Table 3).

The three species exhibited a pattern of very low dispersal rates along most branches with a few outliers corresponding to branches with high dispersal rates (white squares in Fig. 4). Interquartile distances of dispersal rates were similar among species (1.91, 2.13 and 2.63 km/RTU for *P.* ca1, *P.* st1 and *P.* sl1, respectively), but the maximum value, number, and time distribution of outliers differed markedly between species. The highest dispersal rate was approximately 14 and 6 times higher in *P.* ca1 (i.e. 141 km/RTU) than in *P.* st1 (9.9 km/RTU) and *P.* sl1 (24.5 km /RTU), respectively. The 13 outliers in *P.* ca1 appeared during the second half of species life and the two of them showing by far the highest dispersal rates during the last quarter. In contrast, *P.* st1 and *P.* sl1 had only 3 and 5 outliers, respectively. They were distributed in a relatively even manner along species life, although *P.* sl1 lacked any outliers in the last quarter of its life.

Table 3 Results of the spatial diffusion models selection showing Bayes factor (2 ln BF) and selected model for each species. S.E. is for smoothed estimate using 1000 bootstrap replicates.

Species	Model 1	Model 0		Selected model*	Marginal likelihood (+/- S.E.)
		lognormal	Brownian		
P. cavaticus (ca1)	Gamma	104.02	540.82	Gamma (very strong)	-3752 +/- 0.95
	lognormal		436.80		-3804 +/- 0.82
	Brownian				-4023 +/- 0.62
P. strouhali (st1)	Gamma	4.04	118.12	Gamma (positive)	-3489 +/0.63
	lognormal		114.08		-3491 +/- 0.52
	Brownian				-3548 +/- 0.52
P. slavus (sl1)	Gamma	36.54	114.04	Gamma (very strong)	-2670 +/- 0.52
	lognormal		77.5		-2688 +/- 0.49
	Brownian				-2727 +/- 0.39

*Best model and in parenthesis evidence against the second best model.

Range dynamics

Range expansion of the three species during four relative time periods is shown in figure 5. The recent time period (i.e. 75-100% since the oldest species MRCA) roughly corresponds to the Pleistocene since the best absolute age estimates provided by the most probable chronogram for *P.* ca1, *P.* st1 and *P.* sl1 MRCAs were 10.8 (HPD-95%: 4.9-19.2), 10.5 (4.5-18.9), and 5.8 (2.5-10.9) myr, respectively. *P.* ca1 colonized a major proportion (i.e. > 60%) of its present-day distribution area in recent times by dispersing northward in Germany, Belgium and UK, and southward along the French Rhône River (see red lines in Fig. 5). Recent range expansion was more restricted for the two other species. *P.* sl1 recently colonized the Rhine catchment and expanded its range limits within the Danube catchment to the foothills of Austrian Alps. Recent expansion in *P.* st1 was even more limited as this species expanded marginally its range limits into three river catchments (the Rhône, Rhine and Danube River) that it had formerly colonized. Despite singularities of species range dynamics, recent range expansion in all species seemed to have occurred from multiple refugia located at the northern and western periphery of the Alps.

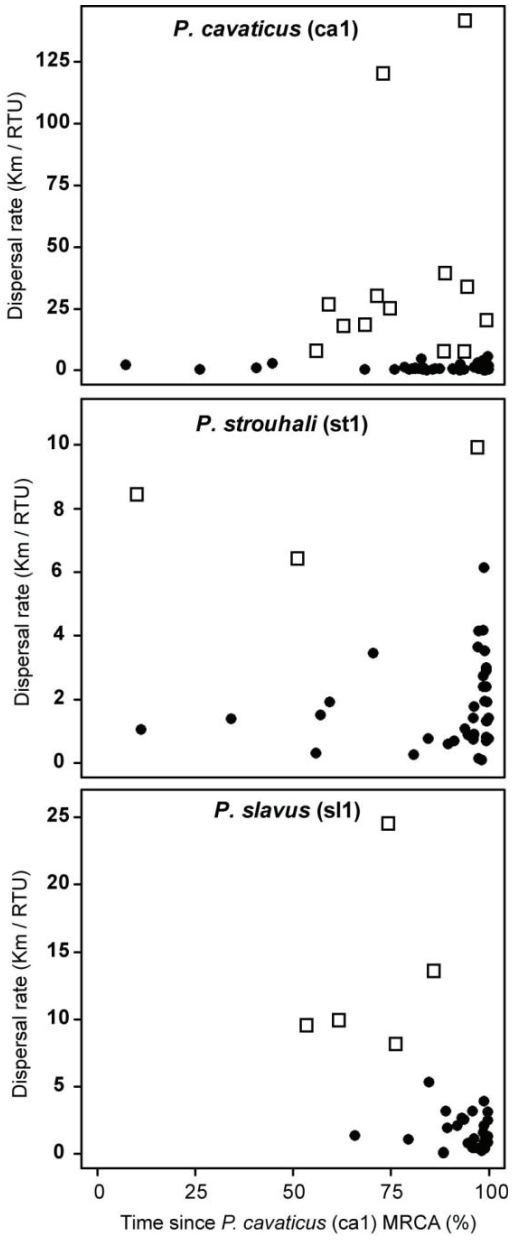


Figure 4: Variation in dispersal rates over the course of a species' evolution. For each branch, dispersal rate is plotted against the mean age of the branch. Relative time units (RTU) correspond to the time duration of a branch divided by the age of the *P. cavaticus* (ca1) most recent common ancestor (MRCA). White squares are defined as outliers defined as branches with dispersal rates higher than 1.5 times the interquartile range (boxplot convention).

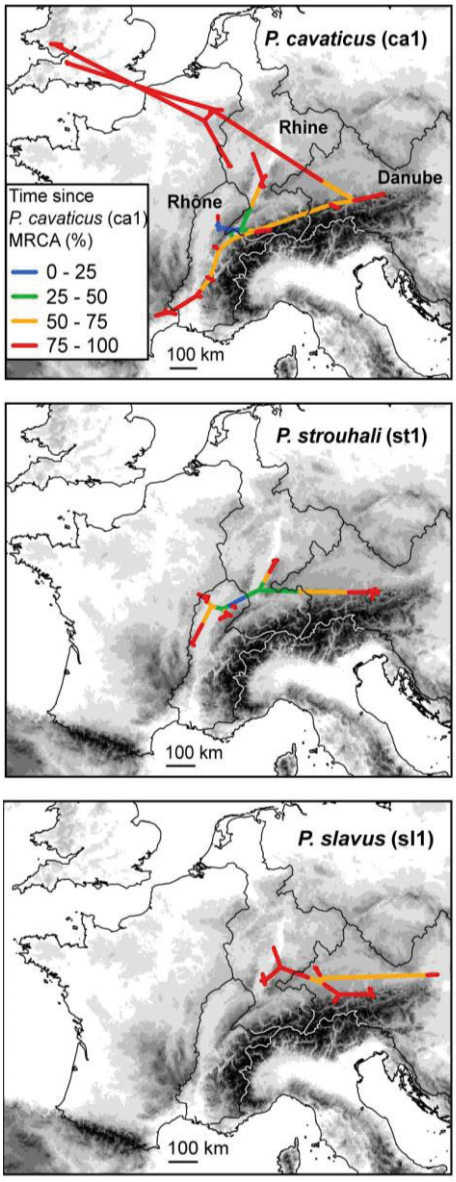


Figure 5: Range dynamics of the three widely distributed cryptic species. MCC tree branches are projected on elevation map and colors represent the relative age of branches (BEAST analysis based on COI, 16S and 28S genes with a linked topology). Time is expressed as percentage of the age of *P. cavaticus* (ca1) most recent common ancestor (MRCA).

Discussion

Historical range dynamics of the five widely-distributed groundwater isopods examined in this study did not support a scenario of post-glacial colonization from distant refugia. Our first prediction that some cryptic species would retain large geographic ranges was partly supported since three cryptic species – *P. cavaticus* (ca1) *P. strouhali* (st1), and *P. slavus* (sl1) - had a geographic range as large as the nominal morphospecies. Phylogeographic spatial diffusion models supported our second prediction by revealing considerable spatiotemporal heterogeneity in dispersal rates over the evolutionary history of the three most widely-distributed cryptic species. Yet, only one of these species, *P. cavaticus* (ca1), experienced a recent, presumably post-glacial, range expansion. Our third prediction that post-glacial colonization might have occurred from distant refugia was not supported since the three species survived cold Pleistocene climates in refugia located at the northern margins of the Alps and foothills of the Jura Mountains.

Cryptic diversity and the size of geographic range

Despite ecological similarities and phylogenetic proximities among the five groundwater isopods, our study revealed unexpected patterns of cryptic diversity with variable consequences on the size of geographic ranges. Trontelj and collaborators (2009) suggested that groundwater species showing range sizes over 200 km were most likely an assemblage of cryptic species with much smaller geographic ranges. Here, we showed three exceptions to this expectation as one cryptic species in each of the three most-widely distributed *Proasellus* morphospecies retained geographic ranges > 660 km, thereby providing evidence of dispersal. Contrary to the assumption that cryptic speciation might be caused by isolation by distance, we did not observe that widely-distributed morphospecies contained more cryptic species than more narrowly-distributed species such as *P. walteri* and *P. synaselloides*. Spatial distribution of cryptic species within morphospecies was also inconsistent with isolation by distance since cryptic species were spatially aggregated. While waiting for further evidence using a larger set of taxa, our results suggest that the spatial distribution of cryptic diversity likely reinforces the pattern of increasing species range size with latitude, documented earlier in distribution studies of groundwater taxa (Malard *et al*. 2009; Stoch & Galassi 2010). This finding adds support to the orbitally forced species' range dynamics theory (Dynesius & Jansson 2000), according to which stronger climatic oscillations at northern latitudes selected for higher dispersal ability and generalism.

The occurrence and highly heterogeneous distribution of cryptic species makes the assessment of groundwater biodiversity an even more challenging issue (Dole-Olivier *et al.* 2009; Trontelj *et al.* 2009). Omitting even a small region such as the Jura Mountains would have led to a strong underestimation of species richness and taxonomic validation of *P. cavaticus* or *P. strouhali*. Morphological taxonomy alone was unable to measure two-thirds of *Proasellus* species richness included in this study. We also found no relationships between cryptic species and formally-recognized subspecies. While calling for an in-depth taxonomic revision (see Table S1) and despite partial agreement between *28S* and *COI* delineations, our results based on *COI* are coherent with recent findings reporting that most groundwater Aselloidea species (i.e. the super-family to which *Proasellus* belongs) contain cryptic species (Morvan *et al*. 2013). This high prevalence of cryptic species is not restricted to the Aselloidea since similar findings were reported for several groundwater taxa (Lefebure *et al.* 2006a*,* 2007; Zakšek *et al.* 2009; Trontelj *et al.* 2009). Together, these studies reinforce the ever increasing view that habitats with reduced environmental heterogeneity such as groundwaters may promote morphological stasis or convergence (Bickford *et al.* 2007).

Spatiotemporal heterogeneity of dispersal

Bayesian phylogeographic diffusion models revealed bursts of dispersal in the three species, which were embedded in a common pattern of weak dispersal. Indeed, dispersal rates appeared to be low along most branches but each species experienced rare events of fast dispersal, which were particularly pronounced in *P. cavaticus*. Our results revive former scenarios suggesting that short-term dispersal windows might have been instrumental in shaping the present-day distribution of groundwater taxa (Magniez 1981). For example, Henry (1976) proposed that *P. cavaticus* had colonized Great Britain during late glacial maximum when north European rivers including the Rhine, Meuse, Weser and Thames Rivers were connected (Bridgland *et al.* 1997). A similar scenario implying late glacial dispersal along the Rhône River from Alpine refugia was proposed to explain the occurrence of *P. cavaticus* in southern France (Henry 1976). During short-time environmental windows of increased habitat connectivity such as periods of intense post-glacial sediment deposition, surface rivers or their permeable subsurface alluvia (i.e. the hyporheic zone) might have provided suitable pathways for dispersal (Ward & Palmer 1994; Buhay & Crandall 2005; Lefébure *et al.* 2006a). However, these dispersal pathways probably did not operate for long durations either because surface water was subsequently colonized by better competitors or subsurface interstices became clogged by finer sediments (Fišer *et al.* 2010; Busschers *et al.* 2005). Our findings of heterogeneous dispersal rates are consistent with results by Buhay & Crandall (2005), suggesting that the genetic structure of groundwater crayfishes in the southeastern United States resulted from a contiguous surface range expansion during short periods of high water levels followed by periods of isolation in caves.

Bayesian relaxed random walks are a key step towards revealing heterogeneous dispersal rates both within and among taxa (Lemey *et al.* 2010; Pybus *et al.* 2012). Yet, we contend that our phylogeographic inferences face several shortcomings. First, they are essentially based on mitochondrial DNA since the 28S nuclear gene lacks variability at the intraspecific level. Despite numerous attempts, we have not been able yet to develop fast evolving nuclear loci such as EPIC markers for our focal taxa. This methodological roadblock is not specific to the present study since there has been no such marker deposited in the GenBank database for Asellota, the suborder to which *Proasellus* belongs. Additional markers would enable us to check for differences between mitochondrial gene and species trees that can be caused by processes such as introgression, stochastic sorting of ancestral polymorphism, drift, and selection. We acknowledge that these processes might have affected the spatiotemporal dynamics of dispersal rates as reflected by mitochondrial genes, in particular our estimates of time lag in dispersal bursts between species. Second, model selection relies on a harmonic mean estimator of marginal likelihood, a procedure that has recently been criticized (Baele *et al.* 2012). Regardless, the strong support for relaxed random walk models against a Brownian random walk model for each species makes us relatively confident in the robustness of dispersal rate heterogeneity. Third, our date estimates would gain precision by incorporating more recent calibration points, which are not yet available for the taxa of interest.

According to earlier scenarios proposed by Henry (1976), we found that *P. cavaticus* (ca1) experienced a major expansion range in recent times, whereas *P. strouhali* (st1) already had acquired much of its distribution range presumably prior to the Pleistocene. *P. slavus* (sl1) expanded along the Danube River before the Pleistocene but its occurrence in the Rhine River catchment and Danubian alpine tributaries is of recent origin. Mechanisms responsible for phase differences in dispersal between species including chance ("to be there at the right time"), selective environmental constraints or biotic interferences are yet to be elucidated.

Location of refugia

Aside from inferring colonization routes, we identified areas that might have acted as Pleistocene refugia. Refugia are areas of higher survival probability due to benign environmental conditions, which may subsequently act as centers of dispersal. In the context of global change, identification of Pleistocene refugia has gained renewed interest for at least three reasons (Nogués-Bravo 2009; Keppel *et al.* 2012). First, occurrence of cryptic refugia can change dispersal estimates since species that are thought to have dispersed over long distances may have survived in far north local refugia (Provan & Bennett 2008). Second, Pleistocene refugia are priority areas for conservation because they constitute climatically stable areas that also may operate during periods of climate warming (Keppel *et al.* 2012). Third, past refugia may be of use to calibrate predictive models of future refugia (Loarie *et al.* 2008).

In Europe, organisms were classified according to the location of refugia in which they survived cold Pleistocene climates, thereby leading to the recognition of Mediterranean, Arctic-Alpine and continental species (Schmitt 2007). Our study showed that the three most widely-distributed species of *Proasellus* were best classified as continental. We rejected the possibility that southern regions might have served as centers of post-glacial dispersal in *P. slavus*, since all populations south of the Alps belonged to narrowly-distributed cryptic species. Moreover, relaxed random walks inferred that recent range expansion in the two other widely-distributed species arose from populations located at the northern margins of the Alps and foothills of the Jura Mountains. Species delineation methods also revealed that the Jura Mountains harbored narrowly-distributed cryptic species within *P. cavaticus* (*P.* ca2-4) and *P. strouhali* (*P.* st2). Overall, these results suggest that the Jura and Alpine foothills played a major role in maintaining diversity within *Proasellus* by acting both as diversification hotspots and subsequent centers of dispersal. Western margins of the Jura also were shown to play a major role in preserving major haplogroups in the hairy land snail *Trochulus villosus* (Dépraz *et al.* 2008) and in the obligate groundwater amphipod *Niphargus virei* (Lefébure *et al.* 2006a; Foulquier *et al*. 2008). Whereas mountain foothills are increasingly being recognized as major continental refugia, causal mechanisms are not yet fully understood. Species might have simply accumulated north of European mountain ranges because the latter were physical barriers to southward dispersal or because mountain foothills acted as true refugia due to increased water availability or higher habitat heterogeneity along elevation gradients (Dépraz *et al.* 2008; Davies *et al.* 2007).

Linking cryptic diversity, range size and dispersal dynamics

One of the major conundrums in biogeography is to find links between patterns and processes. Indeed, cryptic diversity implies that the range size of a morphospecies does not necessarily reflect its realized dispersion. If as suggested in this study, short-time environmental windows of increased habitat connectivity are instrumental in shaping species distribution, even the range size of a well-defined evolutionary unit is hardly evidence of its intrinsic dispersal capacity. Narrowly-distributed species may simply not have had the opportunity to disperse for multiple reasons including chance, abiotic constraints and competition. The development of statistical phylogeography (Knowles & Maddison 2002), approximate Bayesian computation (Beaumont *et al*. 2002) and more recently relaxed random walks (Lemey *et al.* 2010; Bouckaert *et al.* 2012) has successfully contributed evading the many pitfalls on the way from pattern to process. Our study made one step in this direction by revealing that "cryptic speciation" was not necessarily driven by isolation by distance, species distribution was likely the product of asynchronous waves of dispersal and that postglacial colonization was from multiple refugia located at the northern periphery of the Alps.

Acknowledgements

This work was supported by the European Commission (7th EU Framework Programme, Contract No. 226874, BioFresh), the Agence Nationale de la Recherche (ANR–08–JCJC–0120–01, "DEEP") and the Institut Universitaire de France. We thank the following collaborators for their precious help in collecting individuals of *Proasellus*: B. Amont, C. Capderrey, J. Carter, C. Colson-Proch, M. Creuzé des Châtelliers, D. Danielopol, M.-J. Dole-Olivier, A. Fuchs, J.- P. Gramaglia, P. Grenier, S. Gutjahr, S. Heide, P. Jean, B. Kaufmann, L. Knight, R. Le Pennec, A. Malard, F. Mermillod-Blondin, G. Michel, C. Morvan, R. Niederreiter, V. Prie, B. Sket, F. Stoch, and M. Turjak. We gratefully acknowledge TIDRA (http://www.tidra.org) for providing a significant amount of computing resources. We thank D. Culver and C. Robinson for correcting the English and the associate editor and three anonymous reviewers for their valuable comments on an earlier draft of this manuscript.

References

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko V (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, **29**, 2157–2167.

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Bickford D, Lohman DJ, Sodhi NS *et al.* (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, **22**, 148–155.

Bouckaert R, Lemey P, Dunn M *et al.* (2012) Mapping the origins and expansion of the Indo-European language family. *Science*, **337**, 957–960.

Bridgland DR, Gibbard PL (1997) Quaternary river diversions in the London basin and the eastern English channel. *Géographie physique et Quaternaire*, **51**, 337–346.

Buhay JE, Crandall KA (2005) Subterranean phylogeography of freshwater crayfishes shows extensive gene flow and surprisingly large population sizes. *Molecular Ecology*, **14**, 4259–4273.

Busschers FS, Weerts HJT, Wallinga J *et al.* (2005) Sedimentary architecture and optical dating of Middle and Late Pleistocene Rhine-Meuse deposits – fluvial response to climate change, sea-level fluctuation and glaciation. *Geologie en Mijnbouw*, **84**, 25–41.

Calvignac S, Konecny L, Malard F, Douady CJ (2011) Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). *Mitochondrion*, **11**, 246–254.

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, **17**, 540-552.

Charif D, Lobry JR (2007) Seqinr 1.0-2: A contributed package to the r project for statistical computing devoted to biological sequences retrieval and analysis. *Structural Approaches to Sequence Evolution* (eds Bastolla U, Porto M, Roman HE, Vendruscolo M). p. 207–232. Springer, Berlin Heidelberg.

Cordellier M, Pfenninger M (2009) Inferring the past to predict the future: climate modelling predictions and phylogeography for the freshwater gastropod *Radix balthica* (Pulmonata, Basommatophora). *Molecular Ecology*, **18**, 534–544.

Davies R, Orme CDL, Storch D *et al.* (2007) Topography, energy and the global distribution of bird species richness. *Proceedings of the Royal Society B*, **274**, 1189–1197.

Dépraz A, Cordellier M, Hausser J, Pfenninger M (2008) Postglacial recolonization at a snail's pace (*Trochulus villosus*): confronting competing refugia hypotheses using model selection. *Molecular Ecology*, **17**, 2449–2462.

Dole-Olivier MJ, Castellarini F, Coineau N *et al.* (2009) Towards an optimal sampling strategy to assess groundwater biodiversity: comparison across six European regions. *Freshwater Biology*, **54**, 777–796.

Drummond AJ, Ho SYW, Rawlence N, Rambaut A (2007) A rough guide to BEAST 1.4. Available from http://beast.bio.ed.ac.uk/Main_Page.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Dynesius M, Jansson R (2000) Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 9115–9120.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Ezard T, Fujisawa T, Barraclough T (2009) splits: Species' limits by threshold statistics. R package version 1.0-11/r29. Available from http://r-forge.r-project.org/projects/splits/.

Finston T, Johnson M, Humphreys W, Eberhard S, Halse S (2007) Cryptic speciation in two widespread subterranean amphipod genera reflects historical drainage patterns in an ancient landscape. *Molecular Ecology*, **16**, 355–365.

Fišer C, Coleman CO, Zagmajster M, Zwittnig B, Gerecke R, Sket B (2010) Old museum samples and recent taxonomy: A taxonomic, biogeographic and conservation perspective of the *Niphargus tatrensis* species complex (Crustacea: Amphipoda). *Organisms Diversity & Evolution*, **10**, 5–22.

Foulquier A, Malard F, Lefebure T, Douady CJ, Gibert J (2008) The imprint of Quaternary glaciers on the present-day distribution of the obligate groundwater amphipod *Niphargus virei* (Niphargidae). *Journal of Biogeography*, **35**, 552–564.

Furrer R, Nychka D, Sain S (2012) fields: Tools for spatial data. R package version 6.6.3. Available from http://cran.r-project.org/package=fields.

Gouy M, Guindon S, Gascuel O (2010) SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution*, **27**, 221-224.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307-321.

Henry JP (1976) *Recherches sur les Asellidae hypogés de la lignée cavaticus (Crustacea, Isopoda, Asellota).* PhD thesis, University of Dijon, Dijon.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hof C, Brändle M, Brandl R (2008) Latitudinal variation of diversity in European freshwater animals is not concordant across habitat types. *Global Ecology and Biogeography*, **17**, 539–546.

Holsinger JR (1980) *Stygobromus canadensis*, a new subterranean amphipod crustacean (Crangonyctidae) from Canada, with remarks on Wisconsin refugia. *Canadian Journal of Zoology*,**58**, 290–297.

Kass RE, Raftery AE (1995) Bayes Factors. *Journal of the American Statistical Association*, **90**, 773–795.

Keppel G, Van Niel KP, Wardell-Johnson GW *et al.* (2012) Refugia: identifying and understanding safe havens for biodiversity under climate change. *Global Ecology and Biogeography*, **21**, 393–404.

Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology*, **11**, 2623–2635.

Lefébure T, Stanhope MJ (2007) Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biology*, **8**, R71.

Lefébure T, Douady CJ, Gouy M, Gibert J (2006b) Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution*, **40**, 435–447.

Lefébure T, Douady CJ, Gouy M, Trontelj P, Briolay J, Gibert J (2006a) Phylogeography of a subterranean amphipod reveals cryptic diversity and dynamic evolution in extreme environments. *Molecular Ecology*, **15**, 1797–1806.

Lefébure T, Douady CJ, Malard F, Gibert J (2007) Testing dispersal and cryptic diversity in a widely distributed groundwater amphipod (*Niphargus rhenorhodanensis*). *Molecular Phylogenetics and Evolution*, **42**, 676–686.

Lemey P, Rambaut A, Welch JJ, Suchard MA (2010) Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution*, **27**, 1877–1885.

Loarie SR, Carter BE, Hayhoe K *et al.* (2008) Climate change and the future of California's endemic flora. *PloS One*, **3**, e2502.

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2012) *cluster: cluster analysis basics and extensions*. R package version 1.14.3. Available from http://cran.r-project.org/web/packages/cluster/index.html.

Magniez GJ (1981) Biogeographical and paleobiogeographical problems in Stenasellids (Crustacea Isopoda Asellota of underground waters). *International journal of Speleology*, **11**, 71-81.

Malard F, Boutin C, Camacho A *et al.* (2009) Diversity patterns of stygobiotic crustaceans across multiple spatial scales in Europe. *Freshwater Biology*, **54**, 756–776.

Minin VN, Bloomquist EW, Suchard MA (2008) Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, **25**, 1459–1471.

Mittelbach GG, Schemske DW, Cornell H V *et al.* (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, **10**, 315–331.

Morvan C, Malard F, Paradis E, Lefébure T, Konecny-Dupré L, Douady CJ (2013) Timetree of Aselloidea reveals species diversification dynamics in groundwater. *Systematic Biology*, **62**, 512-522.

Nathan R, Klein E, Robledo-Arnuncio JJ, Revilla E (2012) Dispersal kernels: review. In *Dispersal Ecology and Evolution* (eds Clobert J, Baguette M, Benton TG, Bullock JM), pp.187-210. Oxford University Press, Oxford.

Nogués-Bravo D (2009) Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography*, **18**, 521–531.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**, 289-290.

Pons J, Barraclough TG, Gomez-Zurita J *et al.* (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–609.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution*, **23**, 564–571.

Pybus OG, Suchard MA, Lemey P *et al.* (2012) Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 15066–15071.

Rambaut A, Drummond A (2009) Tracer v1.5. Available from http://beast.bio.ed.ac.uk/Tracer.

R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org.

Ricklefs RE (2004) A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, **7**, 1–15.

Schmitt T (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, **4**, 11.

Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution*, **16**, 608–613.

Stoch F, Galassi DMP (2010) Stygobiotic crustacean species richness: a question of numbers, a matter of scale. *Hydrobiologia*, **653**, 217–234.

Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology and Evolution*, **18**, 1001–1013.

Svenning J, Fløjgaard C, Baselga A (2011) Climate, history and neutrality as drivers of mammal beta diversity in Europe : insights from multiscale deconstruction. *Journal of Animal Ecology*, **80**, 393–402.

Taberlet P, Fumagalli L, Wust-Saucy A, Cossons JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Trontelj P, Douady CJ, Fišer C *et al.* (2009) A molecular test for cryptic diversity in ground water: how large are the ranges of macro-stygobionts? *Freshwater Biology*, **54**, 727–744.

Ward JV, Palmer MA (1994) Distribution patterns of interstitial freshwater meiofauna over a range of spatial scales, with emphasis on alluvial river-aquifer systems. *Hydrobiologia*, **287**, 147–156.

Zakšek V, Sket B, Gottstein S, Franjevic D, Trontelj P (2009) The limits of cryptic diversity in groundwater: phylogeography of the cave shrimp *Troglocaris anophthalmus* (Crustacea: Decapoda: Atyidae). *Molecular Ecology*, **18**, 931–946.

Data accessibility:

DNA sequences: GenBank accessions numbers KC6100479-KC610500, KC610160-KC610269, KC610369-KC610418.

Final DNA sequences assembly for testing the monophyly, delineate species and reconstruct Bayesian phylogeographic spatial diffusion models: Dryad doi:10.5061/dryad.fh472

Site characteristics, sequence data set, molecular data acquisition, and BEAST priors for testing monophyly, delineating species and performing Bayesian phylogeographic spatial diffusion models uploaded as online supplemental material.

Supporting information:

Additional supporting information may be found in the online version of this article.

Table S1 Site characteristics and sequence data set.

Table S2 Molecular data acquisition.

Table S3 BEAST priors for testing monophyly, delineating species and performing Bayesian phylogeographic spatial diffusion models.

Table S4 Monophyly support for the three independent genes trees and genetic divergences between individuals within morphospecies and cryptic species.

Fig. S1 Individual genes trees reconstructed with BEAST.

Fig. S2 Subsets of the maximum clade credibility tree with linked topology showing branching patterns within focal taxa.

Fig. S3 28S genes trees of the 5 focal morphospecies showing the distribution of *28S* haplotypes among *COI* cryptic species.

Fig. S4 Frequency distribution of dispersal rates inferred by a Brownian random walk model and a Gamma relaxed random walk model for the 3 widely-distributed species.

Author contributions

DE, FM and CJD designed, performed the research, and wrote the manuscript, DE and TL analysed the data, and LK acquired the DNA sequences.

Chapitre 4) Interactions entre processus de dispersion et de sélection dans la mise en place d'une aire de distribution

Le chapitre précèdent met clairement en avant le rôle de l'hétérogénéité de la dispersion dans la mise en place des aires de répartition actuelles de certaines espèces. Cette hétérogénéité suggère l'existence de courtes fenêtres temporelles facilitant la dispersion des organismes, possiblement *via* l'ouverture d'importants corridors alluviaux lors des épisodes de débâcle glaciaire. Toutefois, il apparait étonnant chez des organismes ne subissant pas de variation thermique saisonnière de maintenir des capacités de dispersion importantes alors même que la sélection devrait favoriser des adaptations locales lors des phases de stabilité climatique.

Ce chapitre, également sous la forme d'un article, a pour objectif d'évaluer le rôle conjoint des processus de dispersion et de sélection vis-à-vis de la température dans l'établissement de l'aire de répartition d'une espèce soumise aux oscillations climatiques du Pléistocène. Ce travail utilise les résultats d'une approche interdisciplinaire couplant phylogéographie, modélisation de la niche et tests physiologiques en laboratoire afin d'établir un scénario explicatif de l'aire de répartition actuellement très fragmentée d'un isopode aquatique souterrain, dont la quasi-totalité des habitats occupés actuellement était recouverte par les glaciers lors du dernier maximum glaciaire (-20 000 ans).

4.1) Article 5 : Integrating phylogeography, physiology and habitat modelling to explore species range determinants

David Eme[1]*, Florian Malard[1], Céline Colson-Proch[1], Pauline Jean[1], Sébastien Calvignac[2], Lara Konecny-Dupré[1], Frédéric Hervant[1] and Christophe J. Douady[1,3]

[1]Université de Lyon, UMR5023 Ecologie des hydrosystèmes Naturels et Anthropisés, Université Lyon1, ENTPE, CNRS, 69622 Villeurbanne, France, [2]Emerging Zoonoses, Robert Koch Institut, 13353 Berlin, Germany, [3]Institut Universitaire de France, Paris, F-75005, France

Article publié dans *Journal of Biogeography* **41**, 687-699.

ABSTRACT

Aim Disentangling the roles of the interacting processes that shape species' ranges requires independent measurements of dispersal, physiological traits and habitat use. Multifaceted approaches of range determinants are, however, still rare, despite the widespread recognition that correlative modelling approaches alone are not sufficient to understand and predict species' distributions. Here, we combined genetic, distributional and physiological data to reveal the processes that cause the disjunct distribution of the groundwater isopod *Proasellus valdensis* in isolated Alpine mountains previously covered by Pleistocene glaciers.

Location The Alps and Jura Mountains, France.

Methods Phylogenetic/phylogeographical methods based on mitochondrial and nuclear genes were used to test for recent dispersal between mountains. A logistic regression on presence–absence data was performed to quantify variation in the probability of occurrence with temperature. Variation in survival and respiration over a range of temperatures was measured within four populations to test for a causal effect of temperature on species distribution.

Results Despite the disjunct distribution, genetic analyses supported recent dispersal between mountains, as indicated by weak divergence among sequences of cytochrome *c* oxidase subunit I (*COI*), a single haplotype network showing no spatial structuring, and a small proportion of molecular variance distributed between mountains. The probability of occurrence of *P. valdensis* decreased significantly with increasing temperature, although physiological experiments indicated that occurrence in warmer habitats was probably restricted by thermally dependent biotic interactions rather than by temperature itself. All populations maintained a high survival rate over a wide range of temperatures (3–15 °C), with a weak but detectable tendency for local adaptation.

Main conclusions Combining phylogeographical, physiological and habitat modelling methods reveals the interacting processes that drive range dynamics. A broad thermal tolerance helps

P. valdensis to colonize vacant habitats during the onset of glacial melting, but range fragmentation and local adaptation, leading to thermal niche narrowing, proceed during warmer interglacials as biotic interactions progressively intensify.

Keywords Alps, climate oscillations, competition, cryptic diversity, gene flow, Proasellus valdensis, respiration rate, species range dynamics, subterranean environment, thermal tolerance breadth.

INTRODUCTION

Explanations for species' range limits necessarily require a dynamic context, because continuing adaptation and/or dispersal is needed to survive in a constantly changing abiotic and biotic environment (Gaston, 2009; Sexton *et al.*, 2009; Geber, 2011). Nonetheless, adaptation and dispersal may not occur on the same time-scale as do environmental changes. Consequently, the geographical range of a species is not primarily the expression of its ecological niche in space. It may be smaller because of dispersal constraints and biotic interactions (Geber, 2011), or larger because of source–sink dynamics and time-lags to extinction (Pulliam, 1988; Oberdorff *et al.*, 2011). Disentangling the relative importance of multiple processes that drive range dynamics often requires independent evidence from distinct approaches. Phylogeographical approaches focusing on dispersal are often necessary to assess whether range limits have recently expanded, contracted or remained stable over long periods (Moeller *et al.*, 2011). Distribution modelling approaches may be informative of species' responses to the abiotic environment (Elith *et al.*, 2006). Most of these methods, however, model realized range rather than potential range (Guisan *et al.*, 2002; Kearney, 2006), and are thus of limited utility in identifying the underlying natural processes. Comparisons of species–environment correlations with independent measures of species physiological tolerances can reveal the role of history of place, barriers to dispersal and biotic processes that often reduce the distribution of a species to a subset of its potential geographical range (Costa *et al.*, 2008). The recognition that species' geographical ranges should be viewed as the outcome of interacting processes necessarily calls for combining single process-orientated approaches (Gaston *et al.*, 2009; Ricklefs & Jenkins, 2011; Wiens, 2011). Multifaceted approaches that use independent measurements of dispersal, habitat use, and physiological traits of species are, however, still relatively rare (Foulquier *et al.*, 2008; Gebremedhin *et al.*, 2009; Moritz *et al.*, 2012).

In this study, we combine genetic, ecological and physiological data to examine how dispersal and selective processes might have interacted over time to shape the present day distribution of the obligate groundwater isopod *Proasellus valdensis* (Chappuis, 1948). This small isopod (5 mm long) exhibits striking distributional features (Fig. 1). First, its geographical range in the Jura Mountains, pre-Alps and Alps is almost entirely contained within areas covered by ice during the Last Glacial

Maximum (Würm glaciation). Second, *P. valdensis* occurs in isolated limestone mountains where it inhabits the vadose and saturated zones of karst aquifers. Repeated attempts to collect this species from the hyporheic zone of rivers that could act as dispersal corridors between isolated mountains were unsuccessful, suggesting that gene flow no longer occurs between mountains. Third, the species has a large elevational range (from 225 to 2670 m a.s.l.), but occurrence data suggest that it is more frequent at high elevation (> 800 m; Henry, 1976). Its range limits suggest a post-glacial colonization following retreat of the Würm glacier, which implies recent gene flow between isolated mountains, but its disjunct distribution may result from a long – presumably pre-Pleistocene – independent evolution of populations in isolated karst mountains. Such long isolation can lead to distinct species that may not necessarily be distinguishable based on morphological criteria, because morphological stasis or convergence can promote cryptic diversity (Trontelj *et al.*, 2009). Lefébure *et al.* (2007) rejected the post-glacial Alpine colonization model for the subterranean amphipod *Niphargus rhenorhodanensis*, demonstrating that this morphospecies is composed of long-isolated cryptic species, some of which probably survived in nunataks during Pleistocene glacial expansion phases. Although correlation does not imply causation, the elevational distribution of *P. valdensis* suggests it might be a cold-adapted species. This species might be able to maximize its physiological performance at low temperature, a trait that would be equally advantageous for surviving *in situ* during glacial periods and for colonizing unoccupied areas at the onset of glacial melting.

Here, we used a multifaceted approach to reveal the processes that cause the disjunct distribution of *P. valdensis* on isolated Alpine mountains previously covered by Pleistocene glaciers. First, we performed phylogenetic/phylogeographical analyses using DNA sequences of two mitochondrial genes – cytochrome *c* oxidase subunit I (*COI*) and *16S* – and one nuclear gene (*28S*) to explore the likelihood of an old cessation of gene flow. Second, a logistic regression model of presence–absence data was performed to assess the relationship between the probability of occurrence and groundwater temperature. Finally, we measured variation in survival and respiration over a range of temperatures within four populations to test for a causal relationship between temperature and the distribution of *P. valdensis*.

MATERIALS AND METHODS

Phylogenetic and phylogeographical data

Sampling was carried out from 2005 to 2010 to obtain specimens of *P. valdensis* from more than one locality on each mountain (Fig. 1, and see Table S1 in Appendix S1 of the Supporting Information). We analysed a total of 48 individuals of *P. valdensis* from 21 caves and springs spanning the species' complete range. Specimens were placed in 96% ethanol at ambient temperature for transportation back to the laboratory and stored at 4 °C until morphological identification. They were identified using

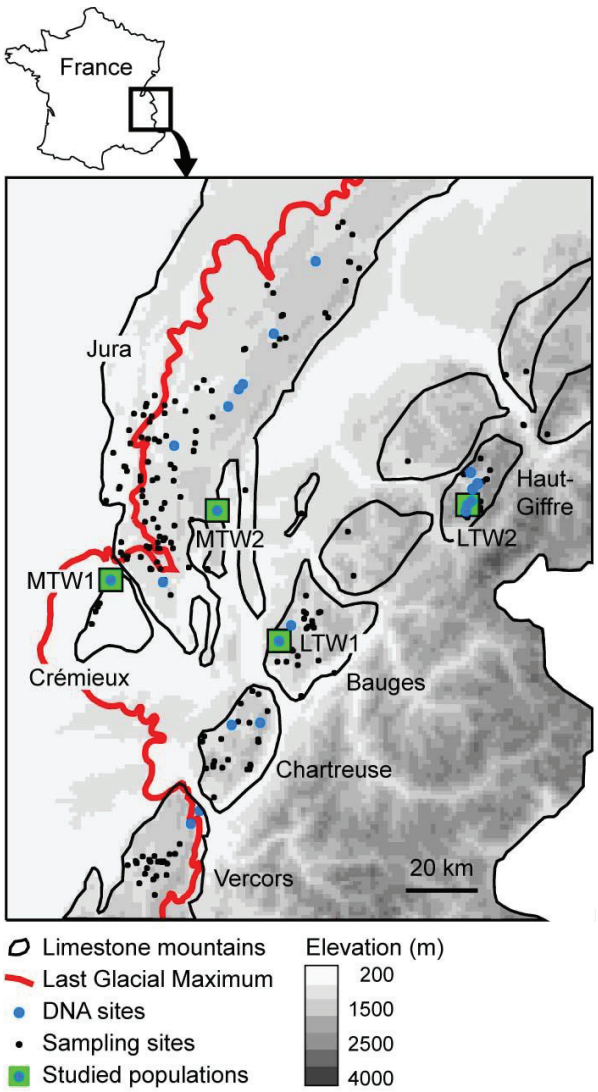


Figure 1: Map depicting the location of Jura and Alps limestone mountains colonized by *Proasellus valdensis* and the distribution of sites for presence–absence data (black dots, $n = 244$), DNA sequences (blue dots, $n = 21$), and physiological measurements (green squares, $n = 4$; MTW, mid-temperature water habitats; LTW, low-temperature water habitats).

species' original diagnoses (Henry, 1976), which is based on the morphology of male copulatory pleopods. Male pleopods were mounted on slides and the remainder of each specimen was stored at −20 °C for molecular analysis. DNA was extracted from three specimens (whenever possible) from each population, following an optimized chloroform DNA extraction protocol (Calvignac *et al.*, 2011). We amplified DNA with primers targeting the mitochondrial *COI* gene, *16S* mitochondrial rDNA gene and *28S* nuclear rDNA gene (Calvignac *et al.*, 2011; Morvan *et al.*, 2013). Polymerase chain reactions (PCRs) followed a previously optimized protocol (Calvignac *et al.*, 2011), although we used a Taq polymerase amount of 0.04 U instead of 0.15 U. PCRs were performed using the following settings: one step of 2 min at 94 °C; 40 (*COI*) or 35 (*16S* and *28S*) cycles of 30 s at 94 °C, 30 s at 48 °C (*COI*), 53 °C (*16S*) or 62 °C (*28S*), 30 s at 72 °C; and one step of 10 min at 72 °C. To prevent the misleading inclusion of nuclear mitochondrial pseudogenes in *COI* and *16S* datasets, we combined three methods for each population (Calvignac *et al.*, 2011): different primer pairs; long-range amplification; and pre-PCR dilution of genomic DNA. Different primer pairs were also used when characterizing *28S*

fragments to detect putative paralogues. Sanger sequencing was performed by service providers (GATC Biotech, Konstanz, Germany; Eurofins MWG Operon, Ebersberg, Germany; BIOFIDAL, Vaulx-en-Velin, France). Chromatograms were visualized using FINCHTV (Geospiza, Seattle, WA, USA). Sequences were submitted to GenBank (accession numbers are given in Table S1 of Appendix S1). Datasets were completed with sequences from 68 Aselloidea species (Morvan *et al.*, 2013). Sequences were then aligned with MUSCLE as implemented in SEAVIEW (Gouy *et al.*, 2010). For each alignment, poorly aligned positions and divergent regions were removed using GBLOCKS (Castresana, 2000). Alignments were declared in GBLOCKS as DNA (*16S* and *28S*) or codon (*COI*) and parameters were set following Castresana's recommendation for less stringent selection.

Presence–absence data and groundwater temperature

Presence–absence data were obtained from the literature, existing distributional databases and specific sampling by the authors for a total of 244 karstic caves and springs. Although we only retained sites where a similar sampling effort had been allocated to search for the target species during the last 15 years, our absence data should be referred to as pseudo-absence. All sites were accurately georeferenced and their elevation was derived from topographic maps. Only 20% of sites were located at elevation > 1000 m because of the difficulty in sampling sites at high elevation. The mean annual groundwater temperature at each site was calculated using a three-step procedure.

First, the mean annual air temperature at each site was derived from a grid coverage of air temperature, with a spatial resolution of 3 km (Rogers, 2003; MétéoSuisse, 2011), using the formula $T_{site} = T_{cell} + [(Elev_{site} - Elev_{cell}) \times -0.0055]$, where T_{site} and T_{cell} are the mean annual air temperature (°C) of the site and grid cell, respectively, $Elev_{site}$ and $Elev_{cell}$ are elevation (m a.s.l.) of the site and grid cell, respectively, and the coefficient −0.0055 is the elevational thermal gradient in °C m^{-1} (Colwell *et al.*, 2008).

Second, we established a linear relationship between mean annual air temperature and mean annual groundwater temperature using groundwater temperature measurements from 379 georeferenced sites located in the study area (Appendix S2).

Third, this relationship was used (1) to transform the grid coverage of air temperature into a grid of groundwater temperature, and (2) to calculate groundwater temperature at the 244 sites for which presence–absence data were available.

Measurements of survival and respiration

The physiological response to temperature variation was measured in two populations of *P. valdensis* living in mid-temperature water (MTW; > 7 °C) and two populations living in low-temperature water (LTW; < 7 °C) (Table 1, Fig. 1). Approximately 100 individuals per population were caught alive and maintained for 6 months in darkness at a constant temperature of 5.5 °C for the

LTW populations and 11 °C for the MTW populations. They were fed *ad libitum* with flake food for aquarium fishes (TetraRubin, Tetra, Melle, Germany). In each population, two groups of 45 individuals were successively used to measure variation in survival rates and respiration rates at temperatures of 1.5, 3.5, 5.5, 9, 11, 13, 15, 17 and 19 °C. These experimental temperatures were reached by increasing or decreasing from the acclimation temperatures (i.e. 5.5 or 11 °C) at a rate of 1 °C change every five days.

Table 1 Location and characteristics of sites in the Alps and Jura Mountains from which populations of *Proasellus valdensis* were sampled for physiological measurements.

Site code	Site Name	Latitude (N)	Longitude (E)	Elevation (m a.s.l.)	Mean annual groundwater temperature (°C)	Annual temperature range (°C)
MTW1	La Balme Cave	45° 51′ 08″	5° 20′ 22″	288	11.6	0.5
MTW2	Huguenots Spring	46° 01′ 52″	5° 46′ 16″	540	10.1	1.5
LTW1	Cavale Cave	45° 39′ 46″	5° 59′ 32″	1370	5.8	0.9
LTW2	Pleureuse Spring	46° 00′ 45″	6° 45′ 31″	1429	4.3	3.3

MTW, mid-temperature water habitat; LTW, low-temperature water habitat.

Individual oxygen consumption was measured in 4.5-mL closed respiration chambers using a microrespiration system (Unisense, Aarhus, Denmark), following Mermillod-Blondin *et al.* (2013). Chambers were placed in thermoregulated baths. A single individual was introduced to each chamber 1 h before starting measurements to avoid respiration peaks caused by animal handling. Isopods were placed on a nylon net with a mesh size of 0.5 mm, and the water in the chamber was continuously mixed with a magnetic stirring rod located below the net. The oxygen microsensor inserted into each chamber was calibrated in air-saturated water (100% oxygen saturation) and Na_2SO_3-saturated water (0% oxygen saturation) prior to each measurement. Oxygen concentration was measured every 30 min for 210 min. The rate of linear decrease of oxygen concentration with time was used to calculate the respiration rate of each individual (μmol O_2 h^{-1} g^{-1} of tissue wet mass), and corrected for the microbial respiration measured in control chambers. Respiration rates were measured for 20–24 individuals per population at each temperature, except at 17 °C and 19 °C where mortality was high.

Data analysis

Monophyly, cryptic diversity and genetic structure

The monophyly of *P. valdensis* was assessed by considering the phylogenetic relationships among 69 Aselloidea morphospecies within a likelihood and Bayesian framework. The three loci were concatenated at the individual level and identical sequences were collapsed using a custom-made Perl script. For haplotype recognition, ambiguities were not considered to be differences between sequences, but gaps were. The best representative sequence for any given haplotype was defined as the longest sequence (containing the fewest ambiguities) using the R software (R Development Core

Team, 2011) and the R package SEQINR (Charif & Lobry, 2007). The most likely topology was inferred in PHYML 3.0 as implemented in SEAVIEW 4.2.12 (Gouy *et al.*, 2010) using a GTR+G+I substitution model, a BIONJ starting tree and the better of nearest-neighbour interchange and subtree pruning and regrafting as the tree search heuristic. Topology robustness was assessed using a bootstrap of 250 pseudo-replicates.

The maximum clade credibility tree (MCCT) was reconstructed with BEAST 1.6.2 (Drummond & Rambaut, 2007) under a GTR+G+I substitution model, using an uncorrelated lognormal Bayesian relaxed molecular clock and a Yule speciation model. Other settings and priors are detailed in Table S2 of Appendix S1. The Markov chain Monte Carlo was run for 200 million generations and sampled every 20,000 generations. Four independent runs were first checked for convergence with TRACER 1.5 (Rambaut & Drummond 2009) and then combined with LOGCOMBINER 1.6.2 (Drummond & Rambaut, 2007), discarding the first 10% of iterations in each run as burn-in. The posterior sample of the trees obtained was summarized by TREEANNOTATOR 1.6.2 (Drummond & Rambaut, 2007) to produce the MCCT and clade posterior probabilities (PP).

We tested for the occurrence of cryptic species within *P. valdensis* using the threshold method defined by Lefébure *et al.* (2006a) and the generalized mixed Yule coalescence (GMYC) model proposed by Pons *et al.* (2006). Lefébure *et al.* (2006a) examined the congruence between molecular divergence and morphological taxonomy within crustaceans using 1500 *COI* sequences from 276 species. These authors proposed that two monophyletic groups diverging by more than 0.16 substitutions per site, as measured by patristic distances, had a strong probability of belonging to different species. The GMYC model is a likelihood method that determines the transition points from species level (pure-birth speciation model) to population level (coalescence). Haplotypes (n = 22 for *P. valdensis*; n = 68 for the other Aselloidea morphospecies) and representative *COI* sequences were defined as described above. For the threshold method, the most likely phylogeny was built using PHYML (Guindon *et al.*, 2010) under a GTR+G+I model of substitution. We used the R packages APE (Paradis *et al.*, 2004) and CLUSTER (Maechler *et al.*, 2012) to compute patristic distances and search for divergent evolutionary units as defined by the threshold of 0.16 substitutions per site. For the GMYC method, a chronogram was reconstructed with BEAST using the best representative *COI* sequence for each haplotype and settings and priors described above. Cryptic species were delineated using the R package SPLITS (Ezard *et al.*, 2009).

Analysis of molecular variance (AMOVA) and haplotype networks were used to explore the population structure of *P. valdensis*. Both approaches were applies only to *COI* and *16S*, because the *28S* locus showed little genetic variation in *P. valdensis*. Site alignment and haplotype recognition were performed as previously described, but with an initial alignment with no outgroup. Genetic variance for *COI* and *16S* was partitioned between and within mountains using one-level AMOVAs performed in ARLEQUIN 3.1 (Excoffier *et al.*, 2005). All individuals belonging to the same mountains were lumped into a single 'population', because there were too few individuals per site to assess

genetic variation within sites. Haplotype divergence was incorporated into the AMOVA by calculating *F*-statistics based on pairwise differences between haplotypes. The statistical significance of variance components was assessed using 20,000 permutations. AMOVA was run first with all mountains, except the Crémieu Mountains, which contained a single site (Fig. 1). We then performed a second run in which the Bauges, Vercors and Chartreuse mountains were combined into a single pre-Alps population to obtain three populations, Jura, Haut-Giffre and the pre-Alps, each containing a similar number of individuals (13–16 individuals for *COI* and 13–15 individuals for *16S*). Haplotype networks were reconstructed using the statistical parsimony approach implemented in TCS (Clement *et al.*, 2000), with a 95% connection limit and gaps treated as missing data.

Probability of occurrence at different temperatures

The relationship between the probability of occurrence of *P. valdensis* and groundwater temperature was assessed using a generalized linear model with logit link and binomial error. Temperature was tested for inclusion in the logistic regression by considering linear and quadratic terms. The reduction in deviance associated with each term was tested for significance at $\alpha = 0.05$ using a chi-square test. The most parsimonious model was selected using the Akaike information criterion. Logistic regression was performed in R. The response curve of *P. valdensis* to temperature was displayed graphically using temperature classes containing an equal number of sampled sites. As recommended by Liu *et al.* (2005), the prevalence – defined as the proportion of presence among sites – was used to determine the temperature above which the species was predicted to be absent by the logistic regression.

Survival and respiration at different temperatures

Repeated-measures analysis of variance (RMANOVA) was used to test for differences in survival and respiration rates among experimental temperatures and between LTW and MTW populations. Temperature was introduced as a repeated-measures factor in the analysis. Post-hoc Tukey's honestly significant difference tests were performed to determine pairwise differences between temperatures. Survival percentages were arcsine-transformed prior to statistical analysis. Significance for all statistical analyses was accepted at $\alpha = 0.05$.

RESULTS

Monophyly, cryptic diversity and genetic structure

The monophyly of *P. valdensis* was strongly supported by both the MCCT (Fig. 2) and the most likely topology (Fig. S1 in Appendix S3). The branche leading to the most recent common ancestor of *P. valdensis* was supported by 1.00 PP and 100% bootstrap. The GMYC model and threshold method did not reveal any cryptic species. The average divergence between any two sequences, as measured

by patristic distances, was 0.05 ± 0.02 substitutions per site. The maximum divergence reached 0.10 substitutions per site, whereas the threshold value above which two *COI* sequences of Crustacea are considered to belong to different species is 0.16. AMOVAs revealed significant genetic variation between and within mountains ($P < 0.001$; Table 2), but only 34.5% of the variation in *COI* and 36.2% in *16S* was distributed between mountains. Even less genetic variation (< 30%) was partitioned between mountains for both markers when the pre-Alps mountains were combined into a single mountain (Table 2).

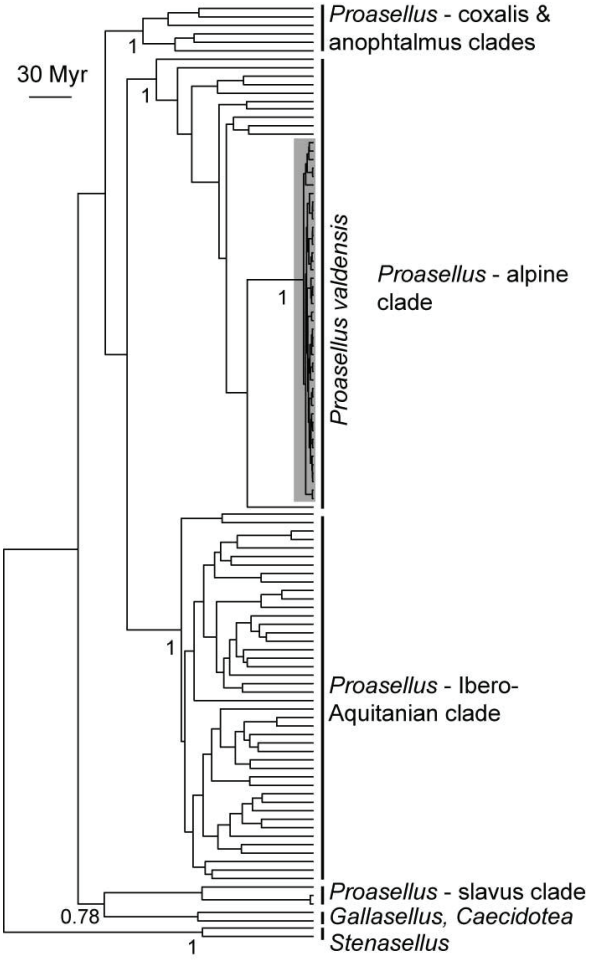


Figure 2: Maximum clade credibility tree of 69 Aselloidea morphospecies in Europe supporting the monophyly of *Proasellus valdensis* (BEAST analysis under an uncorrelated lognormal relaxed molecular clock model). Each terminal branch corresponds to a morphospecies, except for *P. valdensis*. Genera of Aselloidea and main clades within *Proasellus* are shown on the right. Supports for these clades are shown along branches.

All of the 22 *COI* haplotypes were joined in a single network and they did not show any clear geographical distribution pattern (Fig. 3). The inferred ancestral *COI* haplotype was recovered from three isolated mountains – Vercors, Chartreuse and Haut-Giffre – and two of the three other mountains – Jura and Bauges – harboured haplotypes that differed from the ancestral haplotypes by a single mutation (haplotypes 8 and 28). The *16S* haplotype network also revealed no spatial structuring, with one haplotype being shared by populations from Jura and Chartreuse and one by populations from Jura and Haut-Giffre (Fig. S2 in Appendix S3).

Probability of occurrence at different temperatures

Proasellus valdensis occurred at 87 out of 244 sites (Fig. 4). All occurrences but two fell within the limits of formerly glaciated areas. The annual average temperature at sites occupied by the species ranged from 1 to 12 °C. The relationship between the probability of occurrence of *P. valdensis* and groundwater temperature was best described using a linear response curve (coefficient: −0.18, z = −3.33, P = 0.0009) (Fig. 4); the probability of occurrence decreased with increasing temperature and was maximal at temperatures of 2–5 °C. The negative slope of the relationship was highly significant despite the high probability of occurrence at temperature > 11 °C, which was mostly due to the sampling of *P. valdensis* in four hydrologically connected sites close together on Crémieu Mountain. Using a prevalence of 0.36 (87/244) as the occurrence threshold, the logistic regression predicted that *P. valdensis* was absent at temperatures above 8 °C.

Table 2 Analysis of molecular variance between and within mountains for *Proasellus valdensis* in the Alps and Jura Mountains: n = 45 and 42 individuals for cytochrome *c* oxidase subunit I (*COI*) and *16S*, respectively. All variances were significant (P < 0.001).

Source of variation	Five mountains				Three mountains			
	COI		*16S*		*COI*		*16S*	
	Variance	%	Variance	%	Variance	%	Variance	%
Between mountains	1.57	34.5	1.05	36.2	1.03	23.1	0.96	27.7
Within mountains	2.98	65.5	1.85	63.8	3.44	76.9	2.53	72.3
Total	4.55	100	2.90	100	4.47	100	3.49	100

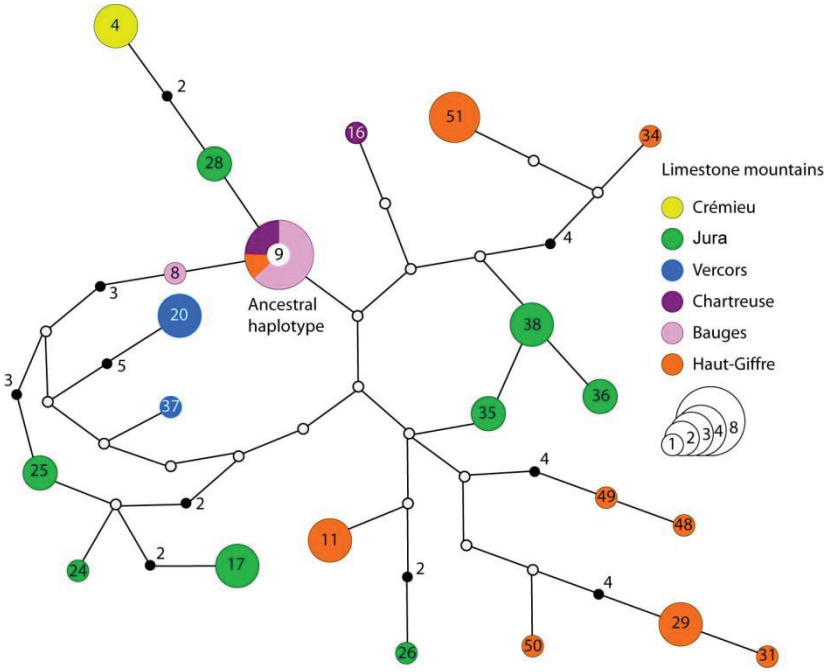


Figure 3: Statistical parsimony haplotype network of cytochrome *c* oxidase subunit I (*COI*) for *Proasellus valdensis* in the Alps and Jura Mountains (n = 48 individuals). The size of each coloured circle is proportional to the number of times that haplotype was sampled. Numbers within circles refer to haplotype codes (see Table S1 in Appendix S1). The size of the coloured sectors within a pie chart is proportional to the numbers of individuals. White and black circles indicate non-sampled or extinct haplotypes. Numbers next to black circles indicate the consecutive number of non-sampled or extinct haplotypes

Survival and respiration at different temperatures

Survival rates under laboratory rearing conditions varied significantly with temperature, but both MTW and LTW populations maintained high survival rates (> 80%) over a wide range of temperatures (Table 3, Fig. 5). Survival rate did not differ significantly between populations (P = 0.869), but there was a significant thermal shift in survival curves between populations (Table 3; interaction between temperature and populations, $P < 0.001$). MTW populations showed no difference in survival between 3.5 and 15 °C, and their survival rates significantly decreased at both ends of this thermal range (post-hoc tests, $P < 0.05$). LTW populations showed no significant mortality at the lowest temperatures and their survival rate significantly decreased at temperatures exceeding 13 °C (post-hoc tests, $P < 0.05$).

Table 3 Results of repeated measures analysis of variance for testing the response (survival and oxygen consumption) of four populations of *Proasellus valdensis* from two different habitats (mid- and low-temperature water) in the Alps and Jura Mountains to temperature variation (n = 90 individuals per population).

Source of variation	d.f.	F	P
Survival			
Habitat	1,2	0.03	0.87
Temperature	8,16	38.66	< 0.001
Habitat × temperature	8,16	9.39	< 0.001
Oxygen consumption			
Habitat	1,2	6.74	0.12
Temperature	8,16	3.85	0.01
Habitat × temperature	8,16	2.00	0.11

Oxygen consumption was significantly influenced by temperature (Table 3, Fig. 6), although both MTW and LTW populations maintained constant respiration at temperatures ranging from 5.5 to 17 °C. MTW populations exhibited higher oxygen consumption rates, but the difference between populations was not significant (P = 0.12). Interactions between the effect of temperature and population were not significant (P = 0.11); both populations decreased their metabolism at low temperatures (i.e. 1.5 and 3.5 °C), but they showed different responses to high temperatures. The oxygen consumption rate of MTW populations decreased when temperature exceeded 17 °C, whereas that of LTW populations increased.

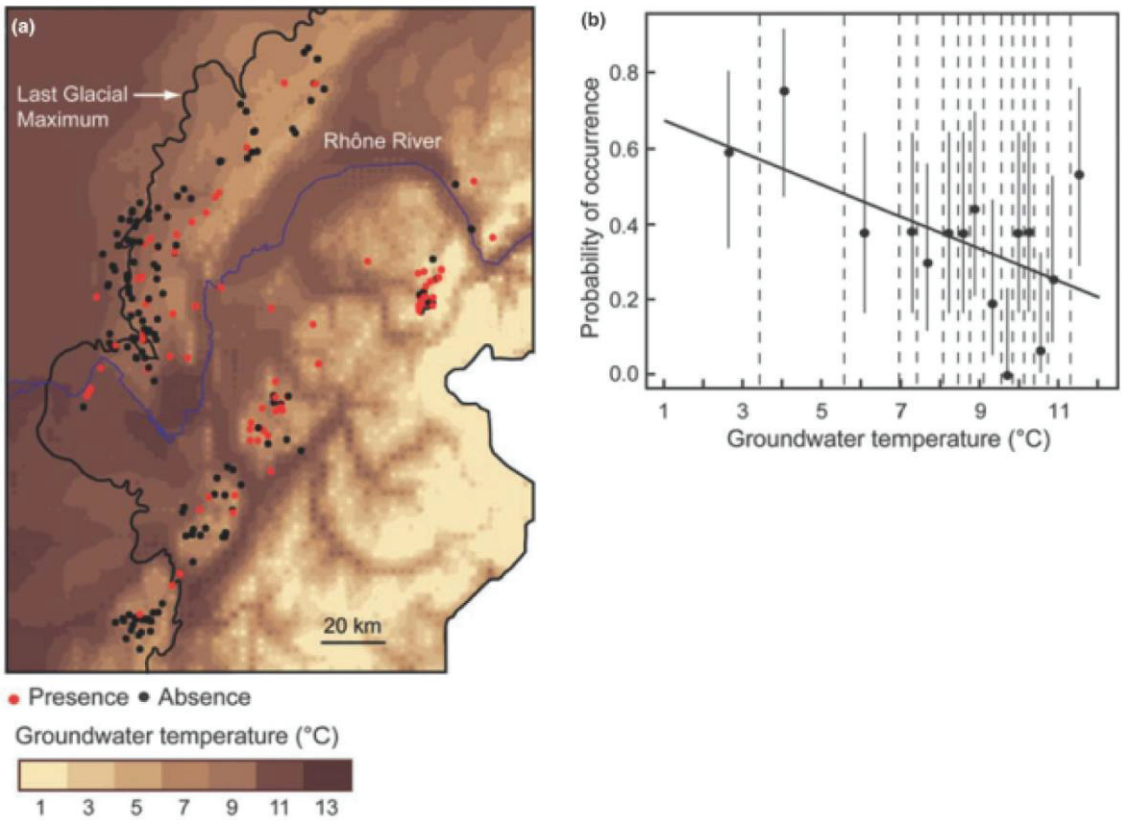


Figure 4: (a) Distribution of *Proasellus valdensis* in the Alps and Jura Mountains and mean annual groundwater temperature. (b) Relationships between the probability of occurrence of *P. valdensis* and mean annual groundwater temperature. Broken lines indicate classes with equal numbers of sampling sites used for calculating probabilities (n = 15 classes, each containing 16–17 sampling sites). Probabilities are plotted at the median of each class. Vertical bars are asymmetrical confidence intervals with continuity correction computed according to Newcombe (1998).

DISCUSSION

Cryptic speciation versus dispersal

Phylogenetic and phylogeographical results were incompatible with a long, independent evolution of populations that led to allopatric speciation in isolated karst mountains. In particular, we found none of the characteristic features expected under a scenario of pre-Pleistocene range fragmentation, including the occurrence of isolated cryptic species, disconnected haplotype networks, and/or a high proportion of molecular variance distributed between mountains (Verovnik *et al.*, 2004; Pauls *et al.*, 2006). The lack of cryptic diversity within *P. valdensis* contrasted with recent studies that revealed the occurrence of cryptic species within subterranean morphospecies (Trontelj *et al.*, 2009; Juan *et al.*, 2010; Morvan *et al.*, 2013). Our findings also contrasted with phylogeographical analyses of the subterranean amphipods *Niphargus virei* and *Niphargus rhenorhodanensis* in the French Jura and Alps that inferred at least two and eight disconnected statistical parsimony networks, respectively (Lefébure *et al.*, 2006b, 2007). Our findings were, however, consistent with earlier mating experiments that showed that populations of *P. valdensis* from the Jura and Alps could produce viable F1 offspring (Henry, 1976). We could not precisely determine the timing of gene flow, especially because both dispersal and stochastic sorting of ancestral polymorphism could contribute to shared genetic variation between mountains. Dispersal between isolated karst mountains might have occurred very recently during Late Glacial Maximum or it might have ceased earlier, during the Pleistocene. In any case, dispersal between mountains implies that *P. valdensis* was once able to colonize surface and/or interstitial habitats of rivers, which constituted the sole migration pathways between isolated karst mountains.

Thermal traits

Our results revealed two distinct and somewhat opposing physiological responses to temperature variation. The first response was that of a generalist species that could maximize its physiological performance, in terms of survival and aerobic respiration, over a wide range of temperature (Huey & Kingsolver, 1989). This broad temperature tolerance was apparent in the four populations, although they experience almost no seasonal variation in temperature in their natural

habitats. For example, LTW population survival from the highest-elevation spring (Pleureuse spring) did not significantly decline until the temperature was above 13 °C, whereas the annual mean and amplitude of spring temperature was only 4.3 °C and 3.3 °C, respectively.

The second response was a weak but detectable tendency between populations to specialize to the thermal conditions encountered in their local habitats. Our experiment was designed to test for differences in physiological response to temperature variation among populations of *P. valdensis* using a classification into two habitat types (MTW and LTW) with two replicate populations from each habitat type. The number of replicates was logistically constrained by the inappropriateness of using fast ramping rates of temperature with groundwater species that experience low thermal variability in their habitats (Mermillod-Blondin *et al.*, 2013). Our survival experiment supported the diagnostic criterion of 'local versus foreign' for local adaptation (Kawecki & Ebert, 2004; Fraser *et al.*, 2011). Populations showed higher survival under temperature conditions resembling those of their local habitats than populations from other habitats.

Our results also revealed a consistent 'home versus away' difference in survival, because each population had a higher survival at temperatures of its own habitat than at other temperatures. Such evidence for local adaptation did not emerge from the comparison of respiration rates, because both LTW and MTW populations were able to maintain a constant rate of oxygen consumption over the same temperature range. The increase in oxygen consumption rate in LTW populations as temperature exceeded 17 °C could reflect an escape behaviour, a response that was not observed in MTW populations. The respiratory response to temperature variation in *P. valdensis* might be constrained by the overriding influence of low food supply on metabolism in most subterranean species (Hervant & Renault, 2002). In particular, the capacity of subterranean organisms to maintain their metabolism as aerobic when confronted to various stresses (i.e. cold, anoxia) instead of using anaerobic metabolic pathways could be driven by the necessity to maintain a high energy yield in a resource-poor environment (Malard & Hervant, 1999; Colson-Proch *et al.*, 2009).

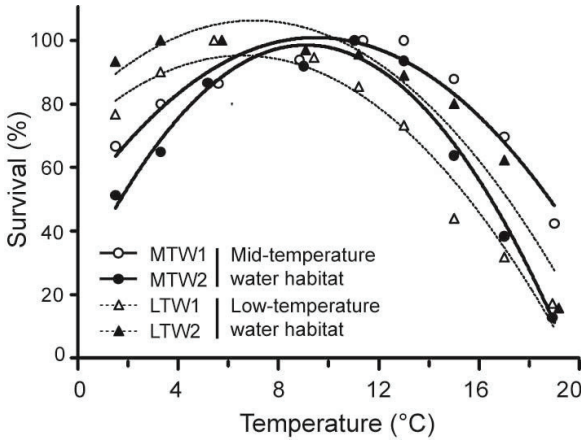


Figure 5: Survival rates measured at nine temperatures (1.5, 3.5, 5.5, 9, 11, 13, 15, 17 and 19 °C) in two populations of *Proasellus valdensis* (n = 90 individuals per population) from mid-temperature water habitats (MTW) and two populations from low-temperature water habitats (LTW) in the Alps and Jura Mountains. Curves were adjusted using a generalized linear model with quadratic terms.

The broad thermal tolerance of *P. valdensis* and the tendency for local adaptation between

LTW and MTW populations might reflect the effect of natural selection acting in opposite directions at two distinct temporal scales. At a temporal scale comprising several Milankovitch oscillations (i.e. several hundred thousand years), climatic changes might have favoured eurythermy by repeatedly placing the species in new thermal conditions. At this scale, groundwater temperature is highly variable because mean annual groundwater temperature in the heterothermic zone closely tracks the air temperature. Theoretical and empirical evidence from a number of studies indicates that Milankovitch climate oscillations selected against specialization and promoted the ability to disperse (Dynesius & Jansson, 2000). Colson-Proch *et al.* (2009) suggested that the broad thermal tolerance of several populations of the subterranean amphipod *Niphargus rhenorhodanensis* in the Jura Mountains – their cold-hardiness in particular – might have been acquired during successive glacial expansion phases in the Pleistocene. At shorter time-scales (i.e. a few tens of thousands of years) corresponding to stable periods between climatic shifts, local adaptation leading to a narrowing thermal niche can proceed because between-site heterogeneity in subterranean temperatures is substantially greater than seasonal variation in temperature. Moreover, lower dispersal during interglacial phases could have restricted gene flow between populations (see below), thereby promoting adaptation to local conditions (Kirkpatrick & Barton, 1997).

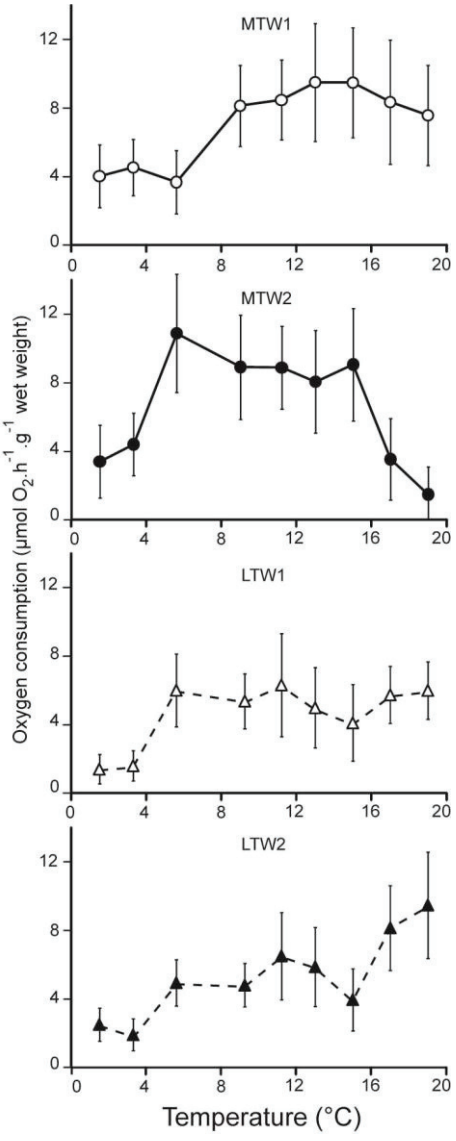


Figure 6 Mean individual oxygen consumption (± standard deviation) measured at nine temperatures (1.5, 3.5, 5.5, 9, 11, 13, 15, 17 and 19 °C) in two populations of *Proasellus valdensis* from mid-temperature water habitats (MTW) and two populations from low-temperature water habitats (LTW) in the Alps and Jura Mountains; n = 20–24 individuals per population at each temperature, except temperatures 17 and 19 °C (n = 14 ± 7) when mortality was high.

Realized range

The fundamental thermal niche of *P. valdensis* should have enabled this species to occupy a wider range, even within the mountains formerly covered by Pleistocene glaciers. Despite a broad thermal tolerance, its probability of occurrence decreased with increasing temperature, indicating that this species most frequently occurred in high-elevation habitats. If, as indicated by physiological and phylogeographical data, *P. valdensis* is a eurythermic species able to disperse recently between mountains, what factors prevented it from occupying low-elevation habitats, where the

temperature overlaps its thermal tolerance? We are unaware of any abiotic factors that would covary with elevation to restrict the occurrence of *P. valdensis* at mid- to low-elevation sites. In particular, inputs of particulate and dissolved organic matter to groundwater should typically be higher in low-elevation, forested karsts than in high-elevation, bare karsts. The most plausible explanation is that the occurrence of *P. valdensis* is restricted by biotic interactions, although further studies are needed to determine the underlying mechanism, including competition for food resources and/or the presence of predators.

We found that populations of *P. valdensis* were restricted to subterranean habitats at low elevation, whereas they were often found to form dense populations in aquatic bryophytes extending several metres from the outlets of high-elevation springs. Fišer *et al.* (2010) suggested that the range fragmentation of *Niphargus tatrensis* species complex during post-Pleistocene climate warming might have resulted from the invasion of competitive species along tributaries of the River Danube. Similarly, our multifaceted approach indicated that *P. valdensis* might have experienced multiple periods of contact and isolation during the Pleistocene. Its broad thermal tolerance might have been advantageous for initially colonizing a large range of habitats either during glacial phases or at the onset of glacial melting. However, the discrepancy between its thermal tolerance and realized distribution suggests that the range of this pioneer species could have secondarily contracted during warmer interglacials as biotic interactions intensified.

Integrating phylogeography, physiology and habitat modelling

Combining genetic, physiological and ecological data suggested that the present-day distribution of *P. valdensis* should be perceived as a snapshot of a highly dynamic range that has been continuously shaped by the interaction between dispersal and abiotic/biotic selective factors during fluctuating Pleistocene climates. Geographical ranges can be extremely dynamic if climatic oscillations over large temporal scales effectively select for generalist strategies that may be described as 'jack of all trades, but master of none' (but see Huey & Hertz, 1984). Such a generalist strategy is advantageous for reaching new habitats in the first place, but initial colonists may be later replaced by better competitors. During short periods that separate climatic shifts, selection may work in the opposite direction, leading to specialization because local adaptation proceeds under conditions of strong spatial variation in selection and low dispersal (Räsänen & Hendry, 2008). Niche narrowing may drive some populations to extinction during the next climatic shift, or local adaptation may be lost as populations are reconnected.

A major uncertainty concerns the temporal scale at which key physiological traits evolve (Roy *et al.*, 2009; Moritz *et al.*, 2012). Time lag in the evolution of traits implies that the physiological performance of a species should be measured independently rather than hypothetically deduced from its present climatic environment. Interestingly, Rapoport's rule (the decline in range size from high to

low latitudes) has stimulated research for more than 10 years (Stevens, 1989; Gaston *et al.*, 1998), whereas Stevens' underlying principle stipulating that the breadth of an organism's thermal tolerance should increase with latitude due to greater seasonal temperature fluctuations has only recently been tested using independent physiological measurements (Calosi *et al.*, 2010; Sunday *et al.*, 2010). If Stevens' principle applies (Sunday *et al.*, 2010), whereas Rapoport's rule does not (Gaston *et al.*, 1998), then species' range limits are unlikely to be in equilibrium with current climate (but see Pigot *et al.*, 2010). A useful way to perceive range dynamics may be to recognize, as stated by Baselga *et al.* (2012), that 'species ranges are simultaneously sensitive to climate and far from in equilibrium with it'.

Times lags in the evolution of traits, dispersal and environmental changes and the interdependency of processes shaping geographical ranges require the integration of independent evidence from evolutionary, phylogenetic, physiological and ecological approaches (Geber, 2011). We contend that such integration will not lead to generality in the causes and dynamic patterns of range limits if it remains specific to particular species or locations. Perhaps the greatest challenge is to extend multifaceted approaches to the study of range limits across multiple species and broad spatial scales. Measuring the breadth of thermal tolerance and the genetic structure of populations between multiple groundwater species that are not exposed to thermal seasonality but differ in their geographical range size would provide a unique opportunity to test for the counteracting effect of maladaptive gene flow on local adaptation (but see Mermillod-Blondin *et al.*, 2013).

ACKNOWLEDGEMENTS

This work was supported by the Agence Nationale de la Recherche (ANR-08-JCJC-0120-01, 'DEEP'), the Institut Universitaire de France, and the European Commission (7th EU Framework Programme, Contract No. 226874, BioFresh). We thank the following collaborators and organizations for their invaluable help in collecting individuals of *P. valdensis*: D. Ariano, Cuves de Sassenage, Grottes de la Balme, B. Hugon, R. Le Pennec, B. Lepretre, E. Mallet, M. Meyssonnier, Réserve Naturelle de Sixt Fer à Cheval and T. Lefébure. We thank D. Fong and C. Robinson for correcting our English and A. Michelle Lawing, the editors and two anonymous referees for their valuable comments on an earlier draft of this manuscript.

REFERENCES

Baselga, A., Lobo, J.M., Svenning, J.-C. & Araújo, M.B. (2012) Global patterns in the shape of species geographical ranges reveal range determinants. *Journal of Biogeography*, **39**, 760–771.

Calosi, P., Bilton, D.T., Spicer, J.I., Votier, S.C. & Atfield, A. (2010) What determines a species' geographical range? Thermal biology and latitudinal range size relationships in European diving beetles (Coleoptera: Dytiscidae). *Journal of Animal Ecology*, **79**, 194–204.

Calvignac, S., Konecny, L., Malard, F. & Douady, C.J. (2011) Preventing the pollution of mitochondrial datasets

with nuclear mitochondrial paralogs (*numts*). *Mitochondrion*, **11**, 246–254.

Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, **17**, 540–552.

Charif, D. & Lobry, J.R. (2007) Seqinr 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. *Structural approaches to sequence evolution* (ed. by U. Bastolla, M. Porto, H.E. Roman and M. Vendruscolo), pp. 207–232. Springer, Berlin Heidelberg.

Clement, M., Posada, D. & Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Colson-Proch, C., Renault, D., Gravot, A., Douady, C.J. & Hervant, F. (2009) Do current environmental conditions explain physiological and metabolic responses of subterranean crustaceans to cold? *Journal of Experimental Biology*, **212**, 1859–1868.

Colwell, R.K., Brehm, G., Cardelús, C.L., Gilman, A.C. & Longino, T. (2008) Global warming, elevational range shifts and lowland biotic attrition in the wet tropics. *Science*, **322**, 258–261.

Costa, G.C., Wolfe, C., Shepard, D.B., Caldwell, J.P. & Vitt, L.J. (2008) Detecting the influence of climatic variables on species distributions: a test using GIS niche-based models along a steep longitudinal environmental gradient. *Journal of Biogeography*, **35**, 637–646.

Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Dynesius, M. & Jansson, R. (2000) Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences USA*, **97**, 9115–9120.

Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47–50.

Ezard, T., Fujisawa, T. & Barraclough, T. (2009) splits: Species' limits by threshold statistics. R package version 1.0-11/r29. Available at: http://r-forge.r-project.org/projects/splits/.

Fišer, C., Coleman, C.O., Zagmajster, M., Zwittnig, B., Gerecke, R. & Sket, B. (2010) Old museum samples and recent taxonomy: a taxonomic, biogeographic and conservation perspective of the *Niphargus tatrensis* species complex (Crustacea: Amphipoda). *Organisms Diversity & Evolution*, **10**, 5–22.

Foulquier, A., Malard, F., Lefébure, T., Douady, C.J. & Gibert, J. (2008) The imprint of Quaternary glaciers on the present-day distribution of the obligate groundwater amphipod *Niphargus virei* (Niphargidae). *Journal of Biogeography*, **35**, 552–564.

Fraser, D.J., Weir, L.K., Bernatchez, L., Hansen, M.M. & Taylor, E.B. (2011) Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity*, **106**, 404–420.

Gaston, K.J. (2009) Geographic range limits of species. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 1391–1393.

Gaston, K.J., Blackburn, T.M. & Spicer, J.I. (1998) Rapoport's rule: time for an epitaph? *Trends in Ecology and Evolution*, **13**, 70–74.

Gaston, K.J., Chown, S.L., Calosi, P., Bernardo, J., Bilton, D.T., Clarke, A., Clusella-Trullas, S., Ghalambor,

C.K., Konarzewski, M., Peck, L.S., Porter, W.P., Portner, H.O., Rezende, E.L., Schulte, P.M., Spicer, J.I., Stillman, J.H., Terblanche, J.S. & van Kleunen, M. (2009) Macrophysiology: a conceptual reunification. *The American Naturalist*, **174**, 595–612.

Geber, M.A. (2011) Ecological and evolutionary limits to species geographic ranges. *The American Naturalist*, **178**, S1–S5.

Gebremedhin, B., Ficetola, G.F., Naderi, S., Rezaei, H.-R., Maudet, C., Rioux, D., Luikart, G., Flagstad, Ø., Thuiller, W. & Taberlet, P. (2009) Combining genetic and ecological data to assess the conservation status of the endangered Ethiopian walia ibex. *Animal Conservation*, **12**, 89–100.

Gouy, M., Guindon, S. & Gascuel, O. (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, **27**, 221–224.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

Guisan, A., Edwards, T.C. & Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, **157**, 89–100.

Henry, J.-P. (1976) *Recherches sur les Asellidae hypogés de la lignée cavaticus (Crustacea, Isopoda, Asellota).* PhD Thesis, University of Dijon, Dijon.

Hervant, F. & Renault, D. (2002) Long-term fasting and realimentation in hypogean and epigean isopods: a proposed adaptive strategy for groundwater organisms. *Journal of Experimental Biology*, **205**, 2079–2087.

Huey, R.B. & Hertz, P.E. (1984) Is a jack-of-all-temperatures a master of none? *Evolution*, **38**, 441–444.

Huey, R.B. & Kingsolver, J.G. (1989) Evolution of thermal sensitivity of ectotherm performance. *Trends in Ecology and Evolution*, **4**, 131–135.

Juan, C., Guzik, M.T., Jaume, D. & Cooper, S.J.B. (2010) Evolution in caves: Darwin's 'wrecks of ancient life' in the molecular era. *Molecular Ecology*, **19**, 3865–3880.

Kawecki, T.J. & Ebert, D. (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Kearney, M. (2006) Habitat, environment and niche: what are we modelling? *Oikos*, **115**, 186–191.

Kirkpatrick, M. & Barton, N.H. (1997) Evolution of a species' range. *The American Naturalist*, **150**, 1–23.

Lefébure, T., Douady, C.J., Gouy, M. & Gibert, J. (2006a) Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution*, **40**, 435–447.

Lefébure, T., Douady, C.J., Gouy, M., Trontelj, P., Briolay, J. & Gibert, J. (2006b) Phylogeography of a subterranean amphipod reveals cryptic diversity and dynamic evolution in extreme environments. *Molecular Ecology*, **15**, 1797–1806.

Lefébure, T., Douady, C.J., Malard, F. & Gibert, J. (2007) Testing dispersal and cryptic diversity in a widely distributed groundwater amphipod (*Niphargus rhenorhodanensis*). *Molecular Phylogenetics and Evolution*, **42**, 676–686.

Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 2**8**, 385–393.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. (2012) *cluster: cluster analysis basics and extensions*. R package version 1.14.3. Available at:

http://cran.r-project.org/web/packages/cluster/index.html.

Malard, F. & Hervant, F. (1999) Oxygen supply and the adaptations of animals in groundwater. *Freshwater Biology*, **41**, 1–30.

Mermillod-Blondin, F., Lefour, C., Lalouette, L., Renault, D., Malard, F., Simon, L. & Douady, C.J. (2013) Thermal tolerance breadths among groundwater crustaceans living in a thermally constant environment. *Journal of Experimental Biology*, **216**, 1983–1994.

MétéoSuisse (2011) *Office fédéral de météorologie et de climatologie MétéoSuisse*. Available at: http://www.meteosuisse.admin.ch/.

Moeller, D.A., Geber, M.A. & Tiffin, P. (2011) Population genetics and the evolution of geographic range limits in an annual plant. *The American Naturalist*, **178**, S44–S61.

Moritz, C., Langham, G., Kearney, M., Krockenberger, A., VanDerWal, J. & Williams, S. (2012) Integrating phylogeography and physiology reveals divergence of thermal traits between central and peripheral lineages of tropical rainforest lizards. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 1680–1687.

Morvan, C., Malard, F., Paradis, E., Lefébure, T., Konecny-Dupré, L. & Douady, C.J. (2013) Timetree of Aselloidea reveals species diversification dynamics in groundwater. *Systematic Biology*, **62**, 512–522.

Newcombe, R.G. (1998) Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine*, **17**, 857–872.

Oberdorff, T., Tedesco, P.A., Hugueny, B., Leprieur, F., Beauchard, O., Brosse, S., & Dürr, H.H. (2011) Global and regional patterns in riverine fish species richness: a review. *International Journal of Ecology*, **2011**, 967631.

Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289-290.

Pauls, S.U., Lumbsch, H.T. & Haase, P. (2006) Phylogeography of the montane caddisfly *Drusus discolor*: evidence for multiple refugia and periglacial survival. *Molecular Ecology*, **15**, 2153–2169.

Pigot, A.L., Owens, I.P.F. & Orme, C.D.L. (2010) The environmental limits to geographic range expansion in birds. *Ecology Letters*, **13**, 705–715.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D. & Vogler, A.P. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–609.

Pulliam, H.R. (1988) Sources, sinks, and population regulation. *The American Naturalist*, **132**, 652-661.

Rambaut, A. & Drummond, A. (2009) Tracer v1.5. Available at: http://beast.bio.ed.ac.uk/Tracer.

R Development Core Team (2011) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.r-project.org/.

Räsänen, K. & Hendry, A.P. (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters*, **11**, 624–636.

Ricklefs, R.E. & Jenkins, D.G. (2011) Biogeography and ecology: towards the integration of two disciplines.

Philosophical Transactions of the Royal Society B: Biological Sciences, **366**, 2438–2448.

Rogers, C. (2003) *Interpolation des températures moyennes mensuelles en France métropolitaine avec l'extension Geostatistical Analyst d'ArcGis 8*. SIG 2003: la Conférence Francophone ESRI, 1–2 October 2003, Issy les Moulineaux, France. Available at: http://www.esrifrance.fr/sig2003/Communication/crenamstet/crenamstet_esri.htm.

Roy, K., Hunt, G., Jablonski, D., Krug, A.Z. & Valentine, J.W. (2009) A macroevolutionary perspective on species range limits. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 1485–1493.

Sexton, J.P., McIntyre, P.J., Angert, A.L. & Rice, K.J. (2009) Evolution and ecology of species range limits. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 415–436.

Stevens, G.C. (1989) The latitudinal gradient in geographical range: how so many species co-exist in the tropics. *The American Naturalist*, **133**, 240–256.

Sunday, J.M., Bates, A.E. & Dulvy, N.K. (2011) Global analysis of thermal tolerance and latitude in ectotherms. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 1823–1830.

Trontelj, P., Douady, C.J., Fišer, C., Gibert, J., Gorički, S., Lefébure, T., Sket, B. & Zakšek, V. (2009) A molecular test for cryptic diversity in ground water: how large are the ranges of macro-stygobionts? *Freshwater Biology*, **54**, 727–744.

Verovnik, R., Sket, B. & Trontelj, P. (2004) Phylogeography of subterranean and surface populations of water lice *Asellus aquaticus* (Crustacea: Isopoda). *Molecular Ecology*, **13**, 1519–1532.

Wiens, J.J. (2011) The niche, biogeography and species interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 2336–2350.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Sampling sites, morphospecies, accession numbers and priors used in BEAST for assessing the monophyly of *Proasellus valdensis* and testing for the presence of cryptic species.

Appendix S2 Relationship between mean annual air temperature and mean annual groundwater temperature in the Alps and Jura Mountains, France.

Appendix S3 Most likely topology of 69 Aselloidea morphospecies and *16S* statistical parsimony haplotype network for *Proasellus valdensis*.

BIOSKETCH

David Eme is a doctoral student in the E_3S team (Ecology, Evolution and Subterranean Ecosystems) at the University of Lyon, France. He is using multifaceted approaches involving species–environment relationships, phylogeography and physiology to reveal historical and contemporary determinants of groundwater biodiversity patterns.

Author contributions: D.E., F.M., C.C.-P. and C.J.D. conceived the ideas and overall study; D.E., F.M. and P.J. analysed the data; C.C.-P. and F.H. performed physiological experiments; S.C. and L.K.-D. carried out the molecular work; F.M. and P.J. assembled presence–absence and temperature data; D.E., C.C.-P., P.J., F.M. and C.J.D. collected specimens; all authors contributed significantly to the final version of the manuscript.

Chapitre 5) Synthèse, discussion, perspectives

5.1) Synthèse des résultats

5.1.1) Rappel des objectifs

L'objectif de ce travail était d'identifier et de quantifier l'influence relative des facteurs environnementaux et des processus impliqués dans la distribution spatiale de la biodiversité souterraine en Europe. Pour atteindre cet objectif, j'ai tout d'abord été amené à utiliser une approche macro-écologique pour établir des inférences sur un cortège important d'espèces (articles 2 & 3). Puis, au risque de perdre en généralité, j'ai concentré mes travaux sur un nombre successivement plus restreint d'espèces afin de gagner en précision sur les relations entre facteurs et processus (articles 4 & 5). A travers cette approche, il s'agissait plus spécifiquement de répondre aux questions suivantes:

- Quels sont, en l'absence de saisonnalité thermique prononcée, les facteurs environnementaux actuels et historiques qui façonnent les patrons de richesse spécifique et de taille des aires de répartition des espèces en Europe ?

- Quel rôle a joué la dispersion, plus spécifiquement lors des périodes postglaciaires, sur les patrons d'aire de répartition en Europe ?

- Comment interagissent les processus de sélection et de dispersion au cours du temps dans l'établissement de l'aire de répartition d'une espèce ?

5.1.2) Construction de bases de données

Trois bases de données ont été construites ou amendées dans le cadre de ce travail. L'acquisition de la base de données des occurrences d'espèces est la pièce maîtresse dans l'étude des patrons de la biodiversité souterraine. La base de données EGCD (European Groundwater Crustacean Database) est le fruit d'un travail collaboratif entrepris avec l'aide de spécialistes de différents groupes de crustacés souterrains. La vérification des listes taxonomiques, la compilation et le géo-référencement de 21700 occurrences regroupant 1570 espèces et sous espèces a pris plus de 2 ans. Ma contribution à cette compilation de données a consisté à rassembler, puis à géo-réferencer des occurrences d'espèces pour les groupes des isopodes, des décapodes et des ostracodes. Cette base de données a permis de documenter pour la première fois les patrons de richesse, de taille d'aire de répartition et de béta diversité à une résolution bien inférieure (100×100km) au grain des écorégions ou des pays jusqu'alors utilisé par des travaux d'étendue spatiale similaire (Hof *et al.* 2008 ; Stoch & Galassi, 2010). Le biais d'échantillonnage a été évalué à partir de plusieurs méthodes afin de tester la robustesse du patron de richesse (article 2).

L'analyse des patrons de biodiversité a par ailleurs nécessité de quantifier les variations spatiales de 80 variables environnementales à l'échelle de l'Europe. La liste de l'ensemble des variables acquise au cours de cette thèse est fournie en annexe 1. La construction de cette base de données environnementale est également le fruit d'un travail collectif qui a été mené en collaboration avec Jean-François Cornu (Institut des Science de l'Evolution Montpellier, ISEM) et Florian Malard (Laboratoire d'Ecologie de Hydrosystèmes Fluviaux Naturels et Anthropisés, LEHNA). Ces variables ont été assemblées à la résolution spatiale d'une maille de cellule de 100*100 km à partir de

bases de données préexistantes (Hijmans et al. 2005, Trabucco & Zomer, 2010, GlobCover 2009) ou de cartes (Ehlers et al. 2004 ; Hughes & Woodward, 2008, Commision for the Geological Map of the World). L'hétérogénéité spatiale des habitats souterrains étant une variable clé pour laquelle je ne disposais d'aucun prédicteur satisfaisant, je me suis également investi dans la réalisation de la première carte des habitats souterrains en Europe (article 1). Cette carte est issue de l'interprétation écologique de la carte hydrogéologique de l'Europe (International Hydrogeological Map of Europe, IHME, http://www.bgr.de/app/fishy/ihme1500/), qui a dû être entièrement vectorisée. Deux paramètres clés – la taille des pores et la perméabilité - ont été utilisés afin d'établir une typologie des habitats et de cartographier leur distribution. Ce travail publié dans *Hydrogeology Journal* a reçu la mention « Editor's choice » en 2014 (Voss *et al.* 2014) et il a également été rendu disponible sans restriction d'usage sur le site web du projet Européen BioFresh (http://data.freshwaterbiodiversity.eu/data/shapefiles/). La carte vectorielle des habitats souterrains a ensuite été utilisée (articles 2 et 3) afin de calculer des prédicteurs permettant de quantifier l'étendue et la diversité des habitats et d'évaluer leurs relations aux patrons de biodiversité souterraine en Europe. Elle offre également bien d'autres perspectives d'utilisation notamment en ce qui concerne la désignation d'aires de conservation de la biodiversité aquatique souterraine en Europe.

Enfin, l'étude de la dynamique des aires de distribution des proaselles (article 4) et de la structure génétique de l'espèce *Proasellus valdensis* (article 5) s'appuient sur des séquences d'ADN (*COI*, *16S* et *28S*) qui sont issues d'une base de données moléculaires pour la super-famille des Aselloidea. Cette base de données, gérée et régulièrement amendée par les membres de l'équipe E3S (Ecologie, Evolution, Ecosystèmes souterrains, UMR-CNRS 2023, LEHNA), compte en 2014 près de 3000 séquences pour un total de 147 espèces et sous-espèces morphologiques (Morvan, 2013). Dans le cadre de cette thèse, j'ai également contribué à amender cette base de données en participant à plusieurs campagnes d'échantillonnage, notamment dans la péninsule balkanique, à l'identification morphologique des proaselles collectés lors de ces campagnes, et de façon plus anecdotique, à l'obtention des séquences d'ADN (extraction d'ADN et PCR).

5.1.3) Résultats marquants de ce travail

Sont brièvement résumés ci-dessous les résultats marquants de mon travail. Ces résultats marquants sont ensuite l'objet d'une description plus détaillée puis ils sont considérés simultanément dans un scénario explicatif global qui permet de comprendre les liens entre facteurs et processus qui ont abouti à la distribution actuelle de la biodiversité souterraine en Europe.

5.1.3.1) Apport des crustacés souterrains dans la compréhension des liens entre patrons de biodiversité, facteurs environnementaux et processus.

Processus à l'origine de la règle de Rapoport

Nos résultats montrent qu'en l'absence de saisonnalité thermique marquée, les oscillations climatiques à long terme sont en elles-mêmes suffisantes pour générer une augmentation de la taille moyenne des aires de répartition des espèces avec la latitude (règle de Rapoport) dans le Paléarctique.

Des facteurs distincts à l'origine des patrons de richesse et de taille des aires de répartition

Alors même que de nombreux travaux soulignent les liens intimes entre patrons de richesse spécifique et de taille des aires de répartition des espèces (i.e. la richesse naît de l'intersection des aires de répartition), nos résultats montrent que ces deux types de patrons peuvent être générés par des facteurs environnementaux distincts.

Multi causalité et non stationnarité spatiale

La richesse spécifique est largement sous l'influence de multiples facteurs dont l'influence relative varie dans l'espace.

Fenêtres temporelles de dispersion

L'étude de la dynamique des aires de répartition montre que la dispersion est un processus extrêmement hétérogène qui intervient au cours des temps géologiques lors de courtes fenêtres temporelles. Ces fenêtres de dispersion ne sont pas synchrones entre espèces mais la dispersion au nord de l'Europe semble s'effectuer principalement à partir de refuges situés le long de l'arc alpin.

Echelles temporelles de la sélection

La sélection pourrait favoriser à des échelles de temps distinctes des traits biologiques favorisant la dispersion et l'adaptation locale.

5.1.3.2) Retour sur les résultats marquants

Règle de Rapoport, beta diversité et variabilité climatique à long terme

Une des difficultés en macro-écologie consiste à expliquer le patron latitudinal de la taille moyenne des aires de répartition, connu sous le nom de règle de Rapoport (Stevens, 1989). Je rappelle que trois principaux facteurs ont été proposés pour expliquer les patrons de taille des aires de répartition : la surface / l'hétérogénéité de l'habitat, la saisonnalité climatique et la variabilité climatique à long terme. Toutefois, la difficulté réside dans la dissociation de l'influence relative des deux échelles temporelles de variabilité climatique en raison de leur co-variation quasi systématique dans les milieux de surface (Morueta-Holmes *et al*. 2013 ; Veter *et al.* 2013). Nos résultats, en supportant le patron de la règle de Rapoport chez les crustacés aquatiques souterrains, apportent une première évidence qu'indépendamment de toute variabilité thermique saisonnière, la variabilité climatique à long terme sélectionne probablement des organismes généralistes et mobiles à l'origine d'une augmentation de la taille moyenne des aires de répartition des espèces avec la latitude (Dynesius & Jansson, 2000). Indépendamment de la latitude, une analyse multifactorielle impliquant le climat actuel, la topographie, la disponibilité d'habitats, et la variabilité climatique à long terme a confirmé que ce dernier prédicteur permettait d'expliquer les variations latitudinales de la taille des aires de répartition des espèces.

Les variations géographiques de la béta diversité et de la contribution relative de ces deux composantes (renouvellement spatial des communautés et diversité emboîtée) découlent de la règle de Rapoport. La contribution de la diversité emboîtée à la béta diversité totale augmente avec la latitude indiquant de potentiels effets de recolonisation au nord par un sous-ensemble d'espèces également présentes dans des régions plus méridionales (Baselga, 2010b). De façon prévisible, nos résultats supportent le fort renouvellement spatial des espèces à travers l'Europe en raison de la dominance des

espèces faiblement distribuées. Toutefois, ils permettent également de constater qu'au-delà de 500 km les communautés sont pratiquement entièrement renouvelées.

Multi causalité et non stationnarité

Ce travail a tout d'abord permis de confirmer la présence d'une crête de richesse spécifique en Europe, préalablement détectée sur une étendue plus réduite (Michel *et al.* 2009). Il a surtout adressé pour la première fois le rôle de la multi-causalité et de la non-stationnarité des causes dans l'espace comme explication du patron de richesse. Nous avons testé l'hypothèse selon laquelle le patron de richesse traduisait les effets multiples des variations climatiques à long terme, de la disponibilité des ressources trophiques (énergie productive) et de l'hétérogénéité spatiale. Ce test fait ressortir 4 points principaux :

1) Analysés séparément, climats historiques, énergie productive et hétérogénéité spatiale disposent du même pouvoir explicatif du patron de richesse. Toutefois, lorsque ces trois facteurs sont examinés conjointement, l'énergie productive et l'hétérogénéité prennent davantage d'importance que le climat historique. Le rôle de l'histoire reste masqué par la variation qu'il partage avec les deux autres facteurs.
2) La richesse est d'autant plus élevée que l'hétérogénéité spatiale est importante, et que le niveau d'énergie est élevé. De la même façon, la richesse augmente d'autant plus que l'hétérogénéité spatiale et la diversité d'habitats sont élevées.
3) La non stationnarité spatiale des effets indépendants et partagés par les trois facteurs fournit l'explication la plus probable de la distribution de la biodiversité souterraine en Europe. L'énergie productive et l'hétérogénéité spatiale sont les deux facteurs prépondérants au niveau de la crête de richesse et plus au sud de celle-ci. Le rôle de la variabilité climatique à long terme se confond avec celui des deux autres facteurs au nord ou seulement avec celui de l'énergie productive au sud. Autrement dit, la conjonction faible diversité d'habitats ayant des pores de taille réduite, faible énergie et forte instabilité climatique au cours de l'histoire, expliquerait la présence d'une plus faible richesse au nord. La faible richesse au centre de la péninsule ibérique serait davantage la conséquence d'une aridité prolongée au cours de l'histoire.
4) Cette étude a aussi permis de détecter le manque de pouvoir prédictif de notre modèle dans la péninsule Balkanique qui reste une zone complexe nécessitant de plus amples investigations, notamment en vérifiant la robustesse de l'échantillonnage.

Hétérogénéité spatio-temporelle du processus de dispersion

L'article 2 révèle des patrons de taille d'aire de répartition et de béta diversité qui suggèrent qu'une petite fraction de la communauté aurait pu recoloniser le nord de l'Europe à partir d'une zone centrale. Cette colonisation n'aurait donc pas eu lieu à partir de refuges dans des péninsules situées plus au sud en Europe. Toutefois, le rôle de la dispersion est de plus en plus souvent remis en cause par la découverte de complexes d'espèces cryptiques ayant de petites aires de répartition au sein d'espèces morphologiques largement distribuées. Dans ce contexte, nous souhaitions tester par une approche de phylogéographie comparative le rôle de la dispersion, et plus spécifiquement celui des colonisations postglaciaires, sur les patrons d'aire de répartition en Europe. Cinq résultats principaux émanent de cette étude menée sur cinq morpho-espèces du genre *Proasellus*.

1) Les grandes aires de répartition pour trois des cinq espèces étudiées ne sont pas un artéfact causé par la présence de complexes d'espèces cryptiques. La dispersion peut donc jouer un rôle important sur la distribution des crustacés aquatiques souterrains.
2) Les grandes aires de répartition se distribuent plutôt au nord de l'Europe (au nord de l'arc alpin) dans des zones impactées par les glaciers et surtout par le permafrost du Pléistocène. En revanche, les espèces cryptiques ayant de petites aires de répartition sont cantonnées quasi-exclusivement au sud de l'arc alpin.
3) La vitesse de colonisation est globalement très lente chez toutes les espèces, mais elle peut s'avérer très rapide lors de quelques évènements rares. Les vitesses maximales sont très hétérogènes entre espèces.
4) La dynamique de colonisation, très variable d'une espèce à l'autre, suggère l'existence de phases asynchrones de dispersion lors de courtes fenêtres temporelles. Ces fenêtres sont possiblement offertes par l'ouverture de corridors alluviaux composés de sédiments grossiers apportés par l'eau de fonte lors des débâcles glaciaires. Toutefois, elles ne seraient pas employées de façon systématique par toutes les espèces. Dans notre étude, seule l'espèce *P. cavaticus* aurait dispersé de manière importante lors du Pléistocène, lui permettant ainsi de coloniser plus de 65% de son aire de répartition actuelle. Les deux autres espèces présentant une large aire de répartition ont colonisé des grands bassins hydrographiques lors de périodes plus anciennes.
5) Enfin, cette étude a permis de localiser les refuges glaciaires ayant permis une recolonisation de l'Europe du nord. Ils s'étendent le long de l'arc alpin, plus précisément à l'ouest du Jura jusqu'aux piémonts des Alpes autrichiennes. La bordure ouest du Jura et le sud des Alpes apparaissent comme des zones de diversification.

Echelles temporelles de la sélection

Dans l'article 5, il s'agissait de mieux comprendre les interactions entre processus de dispersion et de sélection au cours du temps dans l'établissement de l'aire de répartition d'une espèce. Plus précisément, il s'agissait d'établir un scénario plausible à l'origine de l'aire de répartition de l'isopode *P. valdensis* actuellement présent sur des massifs karstiques isolés du Jura et du Nord des Alpes. Ces massifs ont été recouverts par les glaciers lors du dernier maximum glaciaire (-20 000 ans). Nous avons couplé les résultats issus d'approches de phylogéographie, de modélisation de niche et de physiologie afin d'évaluer le rôle de la dispersion, et le décalage entre niches thermiques réalisée et fondamentale. Cette étude croisée supporte un scénario où la dispersion aurait permis une colonisation post glaciaire *via* les corridors alluviaux activés lors de la dernière débâcle glaciaire. Les tests physiologiques en laboratoire entrepris sur plusieurs populations échantillonnées à des températures différentes révèlent l'eurythermie de cette espèce (tolérance entre 2 et 15°C). Toutefois, les différentes populations présentent un début d'adaptation locale. *P. valdensis* n'est donc pas une espèce sténotherme d'eau froide qui aurait suivi son préférendum écologique lors de la retraite des glaciers comme le laissait supposer sa distribution altitudinale. La modélisation de sa probabilité d'occurrence sur le terrain en fonction de la température montre un décalage entre la niche thermique réalisée et la niche fondamentale. La diminution de la probabilité d'occurrence sur la marge la plus chaude de sa distribution suggère que d'autres facteurs, certainement biotiques, contraignent sa distribution.

Outre le scénario biogéographique propre à l'espèce *P. valdensis*, rappelé brièvement ci-dessus, je pense que cet article fait ressortir trois points particulièrement intéressants :

1) La sélection peut agir de façon opposée à des échelles de temps distinctes. En effet, sur des pas de temps longs (oscillations climatiques de Milankovitch), la variabilité climatique sélectionnerait des organismes ayant de meilleures capacités de dispersion leur permettant d'occuper des habitats laissés vacants à la suite du retrait des glaces. Au contraire lors des phases plus courtes de stabilité climatique, la sélection favorise la spécialisation des individus vis-à-vis de leur habitat immédiat *via* l'adaptation locale au détriment de leur capacité de dispersion.
2) Les interactions biotiques pourraient avoir un rôle certainement plus important qu'initialement escompté en milieu souterrain en limitant les opportunités de dispersion.
3) D'un point de vue méthodologique, coupler différentes approches pour établir des sources d'évidences variées et indépendantes permet d'affiner considérablement la vraisemblance des scénarios biogéographiques.

5.1.3.3) Scénario global de la distribution de la diversité des crustacés aquatiques souterrains en Europe

Stoch (1995) a été l'un des premiers à replacer le débat sur la compréhension de la distribution de la biodiversité des organismes souterrains dans un cadre théorique général où l'écologie et le rôle des facteurs environnementaux avaient toute leur place. Toutefois, à la différence de Stoch (1995), je propose un scénario explicatif fondé sur des tests d'hypothèses dont les résultats ont été présentés dans la partie précédente.

En Europe du nord (> 46°N), la très faible richesse résulte vraisemblablement de forts taux d'extinction causés par la succession des glaciations du Pléistocène (Fig.11, zone A). Toutefois, certaines populations auraient pu survivre lors de ces épisodes glaciaires dans des refuges localisés le long de l'arc alpin, du Jura, voire des Ardennes (Fig. 11, zone B). La succession de phases glaciaires et interglaciaires (plus d'une vingtaine) au Pléistocène a certainement exercé une pression de sélection favorisant quelques rares organismes généralistes tolérant une large gamme de températures et disposant de meilleures capacités de dispersion (Dynesius & Jansson, 2000). Ces organismes auraient pour certains recolonisé le nord de l'Europe (l'Angleterre via le fleuve Rhin, la Manche et la Tamise, Fig. 11, zone C). L'extinction préférentielle des espèces présentant de petites aires de répartition et la sélection d'espèces mobiles a conduit à la mise en place d'un gradient latitudinal des aires de répartition (règle de Rapoport). La colonisation de la péninsule scandinave par des espèces de crustacés aquatiques souterrains est quasi-inexistante. Dans cette région de l'Europe, il y a inadéquation entre l'augmentation de la demande énergétique liée à la dispersion, les faibles ressources trophiques, et la qualité des habitats représentés essentiellement par des aquifères à faible perméabilité et à petites tailles de pores. Seules de petites espèces énergétiquement moins efficaces peuvent coloniser ces derniers habitats, il est probable que leur fitness soit affectée par l'inadéquation entre les coûts énergétiques liés à la dispersion et la disponibilité des ressources trophiques.

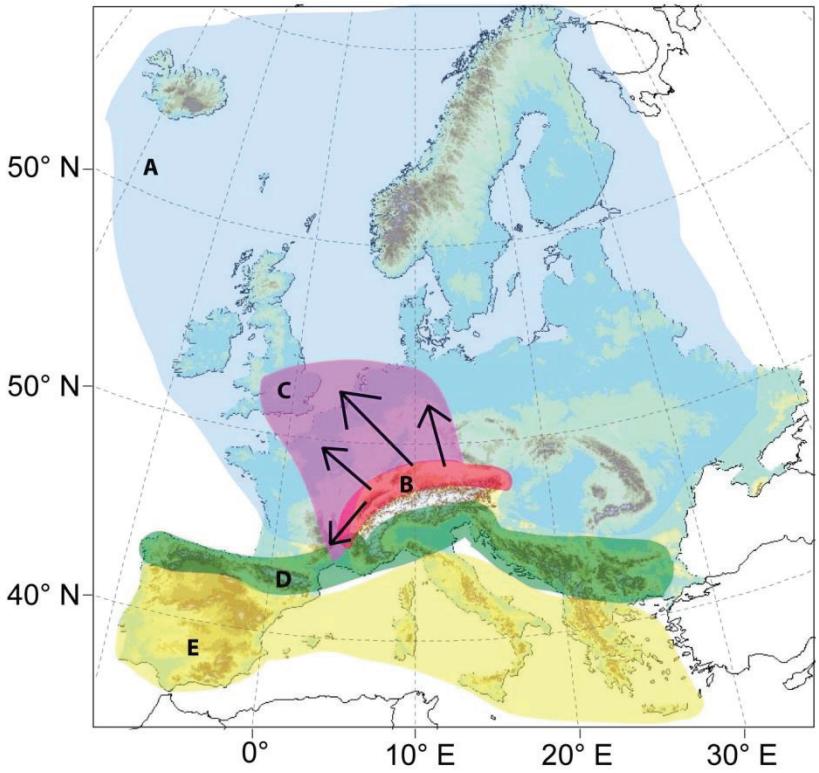


Figure 11 : Scénario explicatif de la distribution des crustacés aquatiques souterrains en Europe. **A)** zone d'extinction préférentielle fortement impactée par les oscillations climatiques du Pléistocène, **B)** zone de refuge glaciaire, **C)** zone de (re)colonisation lors des périodes postglaciaires, **D)** zone montagneuse centre européenne limitant l'extinction et favorisant la spéciation (crête de richesse), **E)** zone de plus forte aridité au cours de l'histoire favorisant l'extinction.

Au centre de l'Europe (Fig. 11 zone D), le long des massifs montagneux, la plus forte hétérogénéité topographique et la grande diversité d'habitats souterrains ont certainement favorisé les spéciations allopatriques et parapatriques (Hewitt, 1996). L'extinction aurait été d'autant plus faible que la stabilité des ressources trophiques au cours du temps aurait permis de maintenir des tailles de population suffisantes (Evans *et al.* 2005). Ces populations ont également eu la possibilité de suivre leurs préférences écologiques en migrant sur des distances très réduites le long du gradient altitudinal (Olhemüller *et al.* 2008). Enfin, le pool régional d'espèces se serait également enrichi d'espèces généralistes à large répartition (mass effect, Stevens, 1989), qui auraient atteint les piémonts montagneux en dispersant le long des multiples affluents connectés aux grands bassins alluviaux.

Au niveau des péninsules du sud de l'Europe (péninsule ibérique, Italie et Grèce, Fig. 11 zone E), la diminution de la richesse pourrait résulter d'une augmentation des taux d'extinction en raison: i) d'une forte aridité au cours de l'histoire qui a limité la disponibilité en eaux et la quantité de ressources trophiques, ii) la forte hétérogénéité spatiale qui tend à limiter les flux de gènes et favoriser la fragmentation des aires de répartition (Gaston & Fuller, 2009 ; Hugueny *et al*. 2011). La combinaison de ces deux facteurs favorise l'entrée des populations dans le vortex d'extinction (Gilpin & Soulé, 1986). La rareté au sein de ces péninsules de grands corridors alluviaux en raison d'une faible influence glaciaire et l'isolement géographique de ces péninsules limitent respectivement la dispersion intra-péninsulaire et l'immigration d'espèces.

5.2) Discussion

5.2.1) Robustesse des patrons de biodiversité

Contribution des espèces largement distribuées

Dans le cadre conceptuel de ce travail (voir paragraphe 1.1.2 « Liens entre métriques »), nous avions évoqué les liens potentiels entre les métriques de biodiversité notamment de richesse, de taille d'aire de répartition et de béta diversité. Ces relations dépendent principalement de la taille des aires de répartition des espèces et de leur agencement dans l'espace (Weiser *et al.* 2007 ; Borregaard & Rahbek, 2010). Ainsi, les espèces ayant de grandes aires disposent généralement d'un poids disproportionné sur l'estimation des facteurs impliqués dans la genèse des patrons de richesse, de tailles d'aire de répartition et de béta diversité. Le nombre d'occurrences par espèce étant généralement proportionnel à la taille de l'aire de répartition, les espèces largement réparties représentent la majorité des occurrences (Jetz & Rahbek, 2002 ; Lennon et al. 2004). Bien que nos inférences soient fondées sur la totalité des espèces souterraines, notamment pour le patron de richesse, je ne pense pas que les quelques espèces largement distribuées puissent affecter profondément ces inférences car la proportion d'espèces ayant de grandes aires (> à 500 km) est très limitée. La surreprésentation des espèces à petites aires de répartition pourrait expliquer les relations relativement ténues entre les patrons de richesse et de taille d'aire de répartition. Elle pourrait aussi expliquer pourquoi les analyses permettent de séparer clairement les facteurs qui influencent chacun des patrons. Toutefois, afin de vérifier ces hypothèses une analyse est actuellement en cours pour d'évaluer si les rôles relatifs de l'histoire, de l'énergie et de l'hétérogénéité spatiale varient fortement selon que sont ou non considérés les 10% d'espèces les plus largement répandues.

Influence des biais d'échantillonnage sur les patrons de biodiversité

Les biais d'échantillonnage ont des conséquences qui sont généralement évaluées uniquement sur les patrons de richesse et non sur ceux de la taille des aires de répartition ou de béta diversité (Gotelli & Colwell, 2001 ; Chao *et al.* 2014). Contraints par des données de présence seulement, nous avons cependant utilisé plusieurs méthodes (d'interpolation ou de raréfaction) qui ont toutes validé la robustesse globale du patron de richesse. Toutefois, il n'est pas à exclure que certaines zones aient pu subir des efforts d'échantillonnage particulièrement déséquilibrés, tels que les Balkans (Culver *et al.* 2006). En effet, le biais d'échantillonnage représente une des hypothèses avancées pour expliquer la faiblesse des prédictions du modèle de richesse dans cette région.

Les répercussions du biais d'échantillonnage sur les patrons d'aires de répartition ou de béta diversité sont plus incertaines et restent très difficiles à appréhender et à corriger. Il est certes possible de corriger statistiquement les effets d'un déficit local d'échantillonnage sur la richesse mais il est, en revanche, impossible d'en évaluer et d'en corriger les effets sur l'aire de répartition et la composition en espèces. Clairement, la fiabilité de l'estimation de la taille moyenne des aires de répartition des espèces dépend directement de la qualité de l'effort d'échantillonnage menée sur la totalité de la zone d'étude (Gaston & Fuller, 2009). La plupart des méthodes ne prennent pas en compte l'incertitude de l'estimation de la taille de l'aire de répartition des espèces mais uniquement notre capacité à détecter la présence d'une espèce à l'échelle locale. Pourtant, ces deux aspects sont cruciaux et intimement liés. Une piste de recherche pourrait consister à tester des méthodes de ré-échantillonnage d'occurrences issues de patrons de diversité simulés. Ainsi, il serait possible d'apprécier la sensibilité des patrons de taille des aires de distribution et de béta-diversité en prenant en compte le biais d'échantillonnage local et ces répercussions sur le biais d'estimation de la taille des aires de répartition

des espèces. A partir de la connaissance issue de ces simulations, il conviendrait de sélectionner les estimateurs de taille d'aire de répartition les moins sensibles au biais d'échantillonnage. En fonction de la résolution spatiale utilisée, il semble raisonnable de penser que les patrons mesurés à partir de métriques d'étendues (EOO) seront moins sensibles que ceux fondés sur la surface occupée (AOO, Gaston & Fuller, 2009).

Influence des biais taxonomiques sur les patrons de biodiversité

Il est maintenant largement reconnu que les fortes contraintes du milieu souterrain favorisent la présence d'une importante diversité cachée (espèces morphologiquement identiques mais génétiquement distinctes) (Lefébure *et al.* 2006a, 2007 ; Trontlej *et al.* 2009 ; Morvan *et al.* 2013). Actuellement, la congruence entre taxonomie morphologique et taxonomie moléculaire n'a pu être évaluée que pour un nombre encore limité d'entités. Toutefois, dans la mesure où la découverte de cette diversité cachée accroît le nombre d'espèces et tend majoritairement à fragmenter les aires de répartition originelle (Trontelj *et al.* 2009), il convient de s'interroger sur la robustesse des patrons actuels de biodiversité et plus particulièrement sur celui de la taille des aires de répartition. Est-ce que la découverte d'espèces cryptiques peut remettre en cause la règle de Rapoport en milieu souterrain ? L'étude moléculaire du troisième chapitre entreprise sur cinq espèces morphologiques du genre *Proasellus* fournit des résultats plus complexes tendant au contraire à renforcer les patrons de richesse et de taille des aires de répartition des espèces. Les méthodes moléculaires d'identification entrainent un découpage des aires de répartition des morpho-espèces plus important au niveau des régions méridionales. Dans ces régions, les espèces cryptiques sont plus nombreuses et possèdent des aires de répartition très étroites souvent limitées à un seul aquifère. Au contraire, dans les régions les plus septentrionales, les entités cryptiques sont moins nombreuses et conservent de larges aires de répartition, d'étendue similaire à celle des morpho-espèces. Ces premiers résultats obtenus sur un nombre restreint de morpho-espèces suggèrent tout de même que la prévalence du cryptisme (probabilité de rencontrer une espèce cryptique) ne serait pas aléatoire dans l'espace mais augmenterait au sud de l'Europe.

Nos résultats sur le genre *Proasellus* sont toutefois contraires à ceux obtenus par une étude moléculaire très récente entreprise sur des espèces d'amphipode du genre *Niphargus* (McInerney *et al.* 2014). Cette étude met en évidence une importante diversité cryptique y compris pour les espèces de *Niphargus* ayant une répartition septentrionale. La découverte de ces entités se traduit par une fragmentation importante des aires de répartition des morpho-espèces (McInerney *et al.* 2014). Cependant, les auteurs utilisent une méthode inhabituelle de délimitation des entités. Cette méthode est particulièrement arbitraire car elle est fondée sur un seuil de divergence extrêmement faible non justifié. Il serait intéressant d'utiliser des méthodes déjà validées (Méthode de seuil, GMYC, Méthode Bayésienne multilocus…) afin de comparer la délimitation des entités et d'évaluer la robustesse du découpage des aires de distribution pour un grand nombre de taxons (Lefébure *et al.* 2006b ; Pons *et al.* 2006 ; Monaghan *et al.* 2009 ; Yang & Rannala, 2010).

Dans ce contexte d'incertitude, il apparait donc primordial de poursuivre les études de phylogéographie sur des espèces à large répartition au nord de l'Europe (par exemple, le copépode harpacticoïde *Parastenocaris glacialis*…). Plus généralement, il faut désormais tester l'hypothèse d'une distribution spatiale homogène de la diversité cachée formulée par Pfenninger & Schwenk, (2007) et tester ensuite la robustesse des patrons de taille des aires de répartition et de richesse établis à partir de données d'occurrence sur des morpho-espèces. Une telle analyse est actuellement menée en collaboration avec des collègues slovènes sur la super-famille des Aselloidea et le genre *Niphargus* dans le cadre du programme PHC Proteus (N°. 31199UM).

5.2.2) Multi causalité : réhabilitation du rôle des ressources trophiques

La vision qui prévalait avant ce travail était celle d'une distribution de la biodiversité aquatique souterraine modelée essentiellement par des facteurs historiques et la fragmentation de l'habitat. Nos résultats supportent une vision plus intégrative où la multi causalité impliquant facteurs historiques, disponibilité des ressources trophiques et hétérogénéité spatiale, constitue l'hypothèse la plus probable. Au cours du prochain paragraphe, nous insisterons plus spécifiquement sur l'importance de la disponibilité des ressources trophiques.

Autant il est reconnu que la richesse et l'abondance des communautés locales aquatiques souterraines sont sous la dépendance de la quantité de ressources trophiques en provenance des écosystèmes de surface, autant le rôle de ce facteur n'avait jamais été réellement envisagé à de larges échelles spatiales (Datry *et al.* 2005 ; Foulquier *et al.* 2011). Seuls Culver *et al.* (2006) ont proposé que la crête de plus forte richesse observée chez les invertébrés souterrains terrestres en Europe et aux Etats-Unis puisse refléter la présence d'une zone à forte productivité énergétique, historiquement épargnée par des épisodes climatiques froids ou des sécheresses. Ce travail de thèse réhabilite le lien fort que la richesse régionale entretient avec l'énergie en général et la quantité de ressources trophiques en particulier. Ce lien entre richesse et énergie mis en évidence chez de très nombreux groupes animaux (Hawkins *et al.* 2003 ; Field *et al.* 2009) a été l'objet d'une série de neuf hypothèses synthétisées par Evans *et al.* (2005). Toutefois, trois principaux mécanismes émanent de cette synthèse. Les deux premiers sont relatifs uniquement au niveau d'énergie ambiante, alors que le dernier dépend de la quantité de ressources trophiques disponible (énergie productive). L'énergie ambiante, souvent approximée par la température, influence directement les organismes *via* deux mécanismes essentiels. Tout d'abord, les niveaux d'énergie ambiante peuvent être plus ou moins compatibles avec la gamme de tolérance physiologique des organismes (Currie *et al.* 2004 ; Evans *et al.* 2005). Dans la mesure où les organismes expriment des performances physiologiques et une « fitness » qui suivent une courbe uni-modale vis-à-vis de la température (Anguilleta et al. 2002), la crête de richesse pourrait représenter la zone où les niveaux d'énergie ambiante sont optimaux pour les crustacés. Toutefois, cette hypothèse reste peu probable car pour des températures équivalentes, la faune souterraine des Alpes est toujours plus diversifiée que celle des régions de haute latitude (21 espèces dans les Alpes à des altitudes > 800 m contre 2 espèces seulement en Scandinavie). Dans un second temps, il a été proposé que l'augmentation de l'énergie ambiante favorisait le processus de spéciation par une accélération des taux de mutations (Rohde 1992 ; Allen *et al.* 2002). Cette accélération des taux de mutation peut provenir directement de l'accélération du métabolisme et de la production de métabolites mutagènes avec la température, où indirectement d'une modification des traits d'histoire de vie qui favorisent les erreurs de copies d'ADN par un accroissement du nombre d'événements méiotiques (Gillooly *et al.* 2002 ; Evans *et al.* 2005). Cette hypothèse a été proposée pour expliquer le gradient latitudinal de richesse où les tropiques apparaissent comme un berceau de diversité (Rohde 1992 ; Jablonsky *et al.* 2006). Mais depuis sa formulation cette hypothèse a reçu un support modéré. En fonction des groupes taxonomiques étudiés, certaines études rapportent des liens entre énergie ambiante et richesse et énergie ambiante et taux de mutation mais ils ne peuvent clairement associer l'augmentation de richesse, aux liens de cause à effet entre énergie ambiante et taux de mutation (Bromham & Cardillo, 2003 ; Davies *et al.* 2004 ; Lanfear *et al.* 2010 ; Goldie *et al.* 2011). Selon toutes vraisemblances, cette hypothèse ne semble pas suffisante à elle seule pour expliquer la crête de richesse des crustacés en Europe, dans la mesure où la richesse devrait être maximale là où la température l'est aussi, c'est-à-dire au sud de l'Europe. Le troisième mécanisme dépend de la quantité de ressources trophiques disponible. Ce facteur largement corrélé à la production primaire et donc ultimement à la dynamique énergie ambiante/disponibilité en eaux, est à la base du

postulat général «more energy, more individuals, more species » (O'Brien, 2006 ; Wright, 1983). D'un point de vue mécanistique, la présence de ressources trophiques suffisamment abondantes permet de limiter le risque d'extinction en maintenant des tailles de populations locales suffisamment importantes, autorisant ainsi l'accueil de d'avantage d'espèces à l'échelle régionale (Evans *et al.* 2005). L'augmentation de la quantité de ressources trophiques peut favoriser la richesse locale en permettant la présence d'espèces spécialisées sur une ressource trophique habituellement rare (« niche position hypothesis ») ou la présence d'espèces spécialisées sur une gamme de ressources très étroite (« niche breath hypothesis ») (Bonn *et al.* 2004 ; Salisbury *et al.* 2012). Elle favorise également l'augmentation du nombre de niveaux trophiques au sein de la chaine alimentaire (« more trophic levels hypothesis ») (Evans *et al.* 2005). Le rôle de l'énergie productive serait particulièrement déterminant dans les zones fortement fragmentées et isolées où une quantité de ressources localement abondantes pourrait permettre de maintenir des tailles de populations suffisantes, y compris en l'absence de migrants (Rahbek & Graves, 2001 ; Ruggiero & Kitzberger, 2004). La localisation de la crête de biodiversité ainsi que les résultats des modèles non stationnaires soutiennent très exactement cette hypothèse.

Les ressources trophiques souterraines particulaires ou dissoutes proviennent de la surface (Gibert & Deharveng, 2002). Toutefois, les communautés microbiennes dans le sol et la zone non saturée en eau (zone vadose) peuvent retenir ou dégrader une partie substantielle de la matière organique avant que celle-ci n'atteigne la nappe phréatique (Datry *et al.* 2005 ; Foulquier *et al.* 2010). Il serait donc intéressant de bénéficier de variables qui permettent de quantifier plus directement les flux de ressources effectivement disponibles en milieu souterrain, tel que le taux de recharge en eau et la concentration en carbone organique dissous (Datry *et al.* 2005). Malheureusement, de tels paramètres ne sont pas actuellement disponibles à de larges échelles spatiales. Leur obtention permettrait également de tester un autre aspect ignoré dans ce travail, celui du rôle de la variabilité saisonnière de la quantité de ressources trophiques. Une forte variabilité constituerait une force de sélection importante qui favoriserait parmi un pool régional d'espèces, celles qui ont par exemple développé des adaptations au jeûne (Hüppop, 2005).

Je terminerais cette partie sur une remarque. Pour de très nombreux groupes d'organismes animaux ou végétaux, l'énergie, qu'elle soit ambiante ou productive est un prédicteur largement corrélé au pic de richesse équatorial (i.e. le gradient latitudinal de richesse ; Hawkins *et al.* 2003 ; Field *et al.* 2009). Etonnamment, les crustacés souterrains pourraient se singulariser car les données d'occurrence à une échelle globale ne font pas apparaître de pic de richesse dans les régions tropicales. Je ne me risquerais pas à émettre des hypothèses explicatives biologiques tant notre incertitude concernant l'effort d'échantillonnage dans la zone intertropicale est grande (Deharveng, 2005), que ce soit en Amérique du sud (Trajano & Bichuette, 2010), en Asie (Brancelj *et al.* 2013) ou plus encore en Afrique (Tuekam kayo *et al.* 2012). Cependant, à l'heure actuelle, il est surprenant de constater que les valeurs de richesse spécifique locale sont plus élevées au niveau de la crête de richesse en Europe que dans les tropiques. Une seule grotte peut communément accueillir plus de 20 taxons dans les régions les plus diversifiées d'Europe (Culver & Sket, 2000 ; Dehaverng *et al.* 2009), ce qui semble être plus rarement le cas au niveau des tropiques alors que les niveaux d'énergie productive en surface sont bien plus élevés (Deharveng, 2005). La forte production primaire en surface ne se traduit-elle pas en milieu souterrain en raison d'une intense activité microbienne dans la zone vadose ? S'agit-il d'un cas hautement débattu dans la littérature, celui d'une valeur théorique maximale au nombre d'espèces (Raboski, 2009 ; Wiens, 2011) ? Ou encore de tous autres mécanismes… Apporter des éléments de réponse nécessite bien évidemment d'intensifier l'effort d'échantillonnage en zone tropicale. Toutefois, il s'avérerait intéressant d'étendre au niveau mondial le consortium européen ayant permis

d'aboutir à la construction de la base EGCD (European Groundwater Crustacean Database, Malard *et al.* 2012), potentiellement *via* la société internationale de biospéléologie.

5.2.3) Importance et hétérogénéité de la dispersion en milieu souterrain

Capacités versus opportunités de dispersion

Les organismes souterrains ont souvent été perçus comme de très mauvais candidats à la dispersion en raison de la petitesse de leur aire de répartition (Gibert *et al.* 2009). Ce travail confirme que la dispersion n'est certes pas l'apanage de la majorité des espèces souterraines mais qu'elle aurait permis pour certaines d'entre elles la colonisation de zones fortement impactées par les oscillations glaciaires du Pléistocène (article 2, 4 & 5). Pour ces quelques espèces, la dynamique d'expansion de l'aire de répartition apparait très hétérogène montrant quelques rares évènements de dispersion très rapide permettant des expansions importantes *via* la colonisation de nouveaux bassins hydrographiques. Ces évènements rapides de colonisation suggèrent que de courtes fenêtres temporelles ont facilité la dispersion. La fonte massive des glaces et du permafrost a généré une érosion considérable (Toucanne *et al.* 2009a,b) permettant l'ouverture de corridors par l'apport massif d'alluvions grossières à forte perméabilité favorables à la dispersion interstitielle (Henry, 1976 ; Ward & Palmer, 1994). De plus, les changements de connexions hydrographiques au cours du temps en raison de la capture de certains cours d'eau ou de mouvements tectoniques auraient autorisé la colonisation de bassins hydrographiques aujourd'hui déconnectés (Schlumberger *et al.* 2001). Ainsi, des connexions entre le Rhin, le Doubs et le Danube ou encore entre la Loire, la Seine et le bassin de la Saône seraient apparues de façon récurrente lors du Pliocène et du Pléistocène (Petit *et al.* 1996). Notons que l'ouverture de connexions apparues exceptionnellement entre bassins hydrographiques couplée à une reconstruction de la dynamique spatio-temporelle de l'aire de répartition permet d'envisager l'utilisation de nouveaux points de calibration pour dater plus précisément les phylogénies, notamment lors des temps récents (moins de 5 millions d'années).

L'asynchronicité des phases de dispersion entre les différentes espèces suggèrent que les capacités intrinsèques à la dispersion ne sont pas suffisantes mais qu'il faudrait en plus bénéficier d'opportunités pour disperser. Ces opportunités de dispersion peuvent dépendre fortement du facteur chance (être au bon endroit au bon moment). Ce facteur chance dépend notamment de la proximité géographique des corridors potentiels, de la fréquence de leurs ouvertures, mais également des interactions biotiques que les populations vont entretenir avec les communautés de surface et souterraine. La fréquence de l'ouverture des corridors dépend certes de la périodicité climatique des cycles de Milankovitch mais aussi des aléas des mouvements tectoniques qui peuvent modifier leur direction ou les rendre inopérants (Toucanne *et al.* 2009a). Les organismes dispersant par les voies interstitielles peuvent être contraints dans leur dispersion par les interférences (compétition, prédation…) avec les communautés épigées. De plus, la compétition pour l'accès aux faibles quantités de ressources trophiques pourrait être déterminante (Gibert & Deharveng, 2002 ; Davies, 2006). Sous ces hypothèses, l'ordre d'arrivée des espèces serait donc crucial dans la réussite de la colonisation (Petermann *et al.* 2010). Le premier arrivant pourrait exploiter les ressources, voire les monopoliser au détriment des autres espèces (Urban *et al.* 2008).

Conflit entre dispersion et adaptation locale.

Dans un milieu où la température est très stable au cours de l'année, le deuxième point clé pouvant limiter la dispersion, outre l'absence d'opportunité, est une adaptation aux conditions locales conduisant à une spécialisation des organismes lors de phases climatiques stables. La réduction de la niche thermique contraint *de facto* les possibilités de dispersion et d'installation dans des zones compatibles avec la valence écologique des organismes (Janzen, 1967). En subissant des oscillations climatiques importantes uniquement sur des pas de temps longs (oscillations glaciaires du cycle de Milankovitch), les organismes sont confrontés à une double échelle temporelle de sélection générant des adaptations à une échelle qui représenteront de potentielles maladaptations à une autre échelle (Räsänen & Hendry, 2008). A long terme, la sélection favorise les organismes généralistes ayant de large valence écologique qui, de ce fait, augmentent leur potentialité de dispersion. En revanche, à plus court terme, la sélection favorise l'adaptation locale et la spécialisation au détriment de la dispersion et de la survie à long terme (Dynesius & Jansson, 2000 ; Jansson & Dynesius, 2002 ; Jocque *et al.* 2010). L'étude menée sur *P. valdensis* illustre parfaitement le cas d'une espèce ayant de bonne capacité de dispersion car disposant d'une large valence thermique. Toutefois, les populations de cette espèce montrent des signes d'adaptation locale qui pourraient représenter une maladaptation lors des prochaines phases glaciaires. Une autre étude, utilisant la taille de l'aire de répartition comme proxi de la dispersion a également soulevé en milieu souterrain ce compromis entre la valence thermique et la dispersion (Mermillod-Blondin *et al.* 2013). Etonnamment, deux espèces endémiques d'un aquifère karstique jurassien présentent des valences thermiques très étroites alors qu'elles ont subi des amplitudes thermiques majeures au cours du Pléistocène. Plusieurs scénarios peuvent être avancés, mais il semble probable que l'adaptation locale ait pu jouer récemment.

Cette dernière hypothèse amène nécessairement à se questionner sur la vitesse d'évolution des traits physiologiques des organismes. Cette question apparait d'autant plus cruciale qu'elle représente pourtant une des clés pour prédire les conséquences des changements climatiques en cours sur le maintien de la biodiversité (Lavergne *et al.* 2010). Toute la difficulté réside dans la compréhension des variations dans le temps et l'espace de la force de la sélection vis-à-vis de la dérive. Notons que cette dernière est elle-même dépendante de la taille efficace des populations et de la migration. Apporter des éléments de réponse concrets nécessite de mener des recherches pluridisciplinaires couplant macrophysiologie (Gaston *et al.* 2009), macroécologie, et phylogéographie afin de générer des résultats généralisables. Plus humblement, une première approche pourrait consister à tester, le long d'un gradient nord-sud sur de multiples espèces non-soumises à une variabilité thermique saisonnière, s'il existe une variation détectable de la sténothermie en réponse aux oscillations climatiques à long terme. La principale difficulté consiste à contrôler des effets de la dispersion quand on sait que la taille moyenne des aires de répartition augmente avec la latitude.

5.2.4) Mesurer les processus plutôt que la diversité

Au cours de cette thèse, j'ai été confronté au compromis entre inférences générales où les processus sont approchés indirectement et inférences plus précises mais peu généralisables. Ce compromis est essentiellement lié au niveau d'organisation biologique à laquelle la métrique quantifiant la diversité biologique est appliquée. L'utilisation de l'espèce morphologique en macro-écologie permet de mener des études généralisables car entreprises sur un grand nombre de taxons. En revanche, l'inférence des processus est largement indirecte *via* l'utilisation de diverses métriques qui, comme nous avons pu le voir, souffrent d'une certaine non-indépendance. La richesse spécifique est une métrique intégrative résultant des processus de spéciation, d'extinction et de dispersion. Le recours aux séquences d'ADN offre la possibilité de quantifier directement les processus de spéciation

et d'extinction, ou tout au moins le taux de diversification (la spéciation moins l'extinction) (Pyron & Burbrink, 2013). Le recours aux nouvelles méthodes de séquençage (Next Generation Sequencing, NGS) devrait permettre de s'approcher plus encore des processus ultimes en mesurant notamment le taux de mutation sur un grand nombre de marqueurs. Cependant, la disponibilité des données moléculaires reste pour le moment limitée à un nombre d'entités biologiques plus restreint, notamment pour des raisons pratiques et historiques. Pourtant, se rapprocher des processus à de larges échelles représente un véritable challenge qui nécessite d'utiliser des métriques de plus en plus proximales, voire de s'en affranchir. Concrètement, je pense que ce challenge peut être relevé en deux étapes.

La première étape consisterait à poursuivre l'acquisition de séquences d'ADN sur un grand nombre d'espèces. Cette acquisition permettrait d'utiliser des métriques plus fines fondées sur une quantification de la divergence évolutive entre entités biologiques, au sein et entre communautés. Ainsi, au lieu de considérer les espèces comme équivalentes, comme le fait la richesse spécifique, évaluer la diversité phylogénétique permet de quantifier la quantité d'évolution présente à un endroit donné ou la dissimilarité évolutive séparant les communautés (Faith, 1992 ; Graham & Fine, 2008 ; Webb *et al.* 2008 ; Cadotte *et al.* 2010). Générer des patrons de diversité phylogénétique permet de distinguer des accélérations de taux de spéciation «*in situ*» et d'étudier plus précisément des phénomènes d'évolution ou de conservation de la niche et de dispersion au sein des communautés (Webb *et al.* 2002 ; Wiens & Donoghue, 2004 ; Graham & Fine, 2008 ; Cavender-Bares *et al.* 2009). L'utilisation de grandes phylogénies (« megaphylogenies », Roquet *et al.* 2013b) datées et couplées à des cartes d'occurrences d'espèces permettrait d'estimer les taux de spéciation/d'extinction et de dispersion qu'il serait possible de cartographier, puis de mettre en relation avec divers facteurs environnementaux actuels et passés (Ronquist & Sanmartin, 2011 ; Condamine *et al.* 2013 ; Eiserhardt *et al.* 2013).

La seconde étape consisterait à se rapprocher des forces évolutives en faisant fi des métriques de diversité. Les nouvelles générations de séquençage (NGS) permettent d'acquérir une quantité d'information génétique considérable, voire la totalité du génome, à des coûts de plus en plus bas et offrent des perspectives inédites pour relever de nouveaux challenges en macro-écologie et macroévolution (Carstens *et al.* 2012 ; McCormack *et al.* 2012). Ainsi, il parait maintenant envisageable même sur des organismes non-modèles d'utiliser des approches de scan génomique pour identifier les loci sous sélection et de tester les relations qu'ils entretiennent avec les différents facteurs environnementaux actuels ou passés (Oleksyk *et al.* 2010; Frichot *et al.* 2013). Il serait également possible de quantifier plus aisément des taux de mutation et des tailles de populations efficaces sur un nombre de marqueurs bien plus élevés (Lanfear *et al.* 2010). Ce gain de puissance statistique permettra potentiellement de découvrir des mécanismes fins liant directement facteurs environnementaux et processus moléculaires dont les répercussions sur la structuration du vivant dans les niveaux d'organisation biologique supérieurs sont jusqu'à présent insoupçonnées.

5.3) Perspectives

Les trois perspectives présentées ci-après traduisent une volonté de se rapprocher des processus impliqués dans la répartition de la biodiversité actuelle. L'enchainement des perspectives reprend la démarche illustrée au cours des deux paragraphes précédents.

5.3.1) Utilisation de la diversité phylogénétique pour affiner la compréhension des liens entre facteurs et processus à l'origine de la crête de richesse

Les résultats en macro-écologie issus de ce travail ont permis d'évaluer l'influence relative des différents facteurs et d'émettre des hypothèses quant au rôle des processus impliqués dans la genèse de la crête de diversité des crustacés dulçaquicoles souterrains en Europe (voir paragraphe 5.1.3.3). Toutefois, l'inférence du rôle des processus reste indirecte et trop imprécise. Cette imprécision est en grande partie liée à l'utilisation de la richesse spécifique comme variable dépendante. La richesse, en plus de considérer les espèces comme équivalentes du point de vue évolutif, représente la résultante des processus de spéciation, d'extinction et de dispersion. L'intégration de la composante génétique en macro-écologie est un champ de recherche émergeant qui offre le double avantage de quantifier les divergences évolutives entre taxons et de réduire la cascade des inférences indirectes pour établir des liens entre facteurs et processus (Emerson *et al.* 2011 ; Davies & Buckley, 2011). De nombreuses métriques quantifiant l'histoire évolutive séparant des taxons ont initialement été développées dans le cadre de la biologie de la conservation et ont été utilisées plus récemment en macro-écologie (Phylogenetic diversity : Faith 1992 ; net relatedness index (NRI) : Webb 2000 ; entropy quadratic (QE) based index : Pavoine *et al.* 2005 ; phylogenetic species variability (PSV) : Helmus *et al.* 2007…). De telles métriques permettent de construire des patrons de diversité phylogénétique dévoilant des zones où les organismes sont plus ou moins apparentés par rapport à un modèle nul (Cavender-Bares *et al.* 2006 ; Forest *et al.* 2007; Hardy & Santerre 2007 ; Morlon *et al.* 2010). Ces déviations de diversité phylogénétique nous renseignent directement sur l'influence relative des processus de spéciation et d'extinction (pour de plus amples détails voir Figure 1 dans Davies *et al.* (2007b)). Par exemple, une forte richesse spécifique associée à une faible diversité phylogénétique signe un processus récent de spéciation *in situ*, souvent associé à des zones présentant un fonctionnement de type « berceaux de diversité ». Au contraire, à richesse spécifique constante, une augmentation de la diversité phylogénétique suggère une diminution des taux d'extinction et signe la présence de zones préservant la diversité à long terme, souvent qualifiées de muséum (Jablonski *et al.* 2006).

Mes résultats issus d'une étude macro-écologique mettent clairement en évidence la présence d'une zone de forte richesse spécifique en Europe qui est associée à des niveaux élevés d'énergie productive et d'hétérogénéité spatiale (Articles 2 et 3). Cette zone résulte-t-elle d'un processus récent de spéciation caractéristique d'un « berceau de diversité » ou d'une extinction réduite caractéristique d'un « muséum » ? Quels sont les facteurs les plus susceptibles de créer de telles zones ? L'objectif de cette perspective est précisément d'identifier les processus à l'origine de cette crête de diversité et de quantifier l'influence relative des facteurs environnementaux sur le patron de diversité phylogénétique. Il s'agira plus précisément de tester si la crête résulte d'un processus de spéciation attendu sous un fonctionnement de type berceau par rapport à l'hypothèse alternative d'une limitation des taux d'extinction supportée par un fonctionnement de type muséum. Sous l'hypothèse du berceau, il est attendu que la forte richesse spécifique au sein de la crête soit associée à une faible diversité phylogénétique. Dans le cas contraire, sous l'hypothèse du muséum, il est attendu qu'à richesse spécifique équivalente, la diversité phylogénétique soit plus forte au niveau de la crête en raison de la diminution des taux d'extinction favorisant le maintien d'une importante diversité au cours du temps.

Le genre *Proasellus* offre un modèle de choix pour tester ces mécanismes. Il présente une distribution *pan* européenne (ouest du Paléarctique) qui recouvre largement la crête de richesse. Ce genre particulièrement riche comptant près de 142 espèces et sous-espèces morphologiques dispose

aussi d'avantages pratiques dans la mesure où l'utilisation de base de données d'occurrences précises acquises au cours de cette thèse (EGCD, Zagmajster *et al.* accepted) permet de connaitre la répartition des espèces et d'estimer la fiabilité de l'échantillonnage. Ajoutons que nous disposons au sein de l'équipe E3S, d'une base de données moléculaires pour la super-famille des Aselloidea incluant près de 3000 séquences de gènes mitochondriaux (*16S* et *COI*) et nucléaires (*28S*), plus de 400 sites et près de 80% des espèces morphologiques du genre *Proasellus*. Cette base de données moléculaire a également permis de développer un cadre phylogénétique robuste pour le genre *Proasellus* (Morvan, 2013).

Dans un premier temps, il conviendra d'utiliser la même grille spatiale (maille de 100*100 km) que pour les articles 2 et 3 ainsi que la base de données d'occurrences des crustacés souterrains (EGCD, Zagmajster *et al.* accepted) afin de définir les cellules où un effort d'échantillonnage de la faune dulçaquicole souterraine peut être considéré comme satisfaisant. A partir de l'arbre phylogénétique du genre (Morvan, 2013), il sera alors possible de calculer les métriques de diversité phylogénétique afin de bénéficier d'une quantité phylogénétique totale par unité spatiale (PD, Faith, 1992) et d'une diversité phylogénétique relative, c'est-à-dire corrigée par la richesse (PDrel, Davies *et al.* 2007b). Des tests de permutations de la localisation des espèces permettront de comparer les patrons observés de DPrel à celui attendu sous un modèle nul de répartition aléatoire des espèces. Ces premières étapes permettront d'établir les premières cartes présentant le patron de la diversité phylogénétique d'organismes aquatiques souterrains à l'échelle de l'Europe afin d'identifier des zones susceptibles de fonctionner comme des berceaux et des muséums de diversité. Dans un second temps, des analyses corrélatives entre facteurs environnementaux passés/actuels et la diversité phylogénétique (PD et PDrel) seront réalisées afin de quantifier l'influence relative de ces différents facteurs et de mettre en évidence ceux qui sont le plus susceptibles de favoriser les taux de spéciation *in situ* ou de diminuer les taux d'extinction au cours du temps.

5.3.2) Muséum ou arche de Noé ? Le recours à la phylobétadiversité

Dans cette deuxième perspective, il s'agira d'affiner notre compréhension des processus et des facteurs impliqués dans la genèse du patron de diversité phylogénétique. Il est tout particulièrement ici question d'explorer le rôle de la dispersion afin de distinguer deux types de fonctionnement : le muséum et l'arche de Noé. Une arche de Noé correspondrait typiquement à une zone maximisant richesse et diversité phylogénétique en raison certes d'un maintien de la diversité au cours du temps mais également en raison d'un fort taux d'immigration. Ce fonctionnement ne peut pas être distingué d'un fonctionnement de type muséum (diminution de l'extinction) à partir d'une seule approche comparative des patrons de richesse spécifique et de diversité phylogénétique. En revanche, la comparaison de la diversité phylogénétique entre entités spatiales permet de quantifier les dissimilarités d'histoire évolutive, ainsi que des temps de divergence moyens entre communautés (Graham & Fine, 2008 ; Webb *et al.* 2008 ; Eiserhardt *et al.* 2013 ; Qian *et al.* 2013). Les indices de phylobétadiversité ont été développés en ce sens (Lozupone & Knight, 2005 ; Bryant *et al.* 2008 ; Graham & Fine, 2008). Les premiers indices de phylobétadiversité développés ne permettaient pas d'évaluer indépendamment le rôle du remplacement spatial et celui des gradients de diversité phylogénétique liés à la différence de richesse entre entités spatiales (Hardy *et al.* 2012 ; Leprieur *et al.* 2012). La décomposition récente des premiers indices de phylobétadiversité offre désormais la possibilité d'évaluer correctement ces deux composantes (Leprieur *et al.* 2012). De ce fait, il est alors possible de distinguer un fonctionnement de type muséum de celui d'arche de Noé. Pour une richesse équivalente, sous un fonctionnement de type muséum, il est attendu que le remplacement spatial de

diversité phylogénétique soit plus élevé entre entités spatiales que sous un fonctionnement de type arche de Noé. Au contraire, la diminution de la proportion de la composante de remplacement spatial phylogénétique est la signature d'un fonctionnement de type arche de Noé car elle traduit la présence des mêmes espèces dans les cellules avoisinantes en raison du processus de dispersion.

Par la suite, il est possible de quantifier l'influence relative des différents facteurs environnementaux sur les composantes de remplacement spatial de la diversité phylogénétique et de gradient de diversité phylogénétique. Plus généralement, cette analyse permet d'estimer conjointement l'importance relative des facteurs sur le processus de diversification (spéciation *versus* extinction) *in situ* et sur ceux de dispersion. Cette perspective envisagée dans le cadre d'une amélioration de notre compréhension des processus et des facteurs à l'origine de la diversité phylogénétique et de la crête de richesse des custacés dulçaquicoles souterrains en Europe peut s'appliquer à tout type d'organismes dans un cadre plus général. A ma connaissance, aucune étude n'a pour l'instant tenté de distinguer les facteurs à l'origine des fonctionnements de type muséum ou d'arche de Noé en utilisant une approche fondée sur une décomposition des composantes de la phylobétadiversité.

D'un point de vue pratique cette analyse pourrait être entreprise sur le genre *Proasellus* en raison des nombreux avantages techniques liés à la disponibilité des données moléculaires et d'occurrences (voir paragraphe précédent pour la disponibilité des données). En ré-utilisant la sélection des entités spatiales présentée en perspective 1, l'arbre phylogénétique du genre combiné à la base de données d'occurrences géo-référencées permettra de calculer les métriques issues de la décomposition d'un indice de phylobétadiversité Unifrac (Lozupone & Knight, 2005 ; Leprieur *et al.* 2012). Cet indice offre l'avantage d'être basé sur la formulation de l'indice de dissimilarité de Jaccard : sa décomposition sera directement comparable à celle de la béta-diversité réalisé dans l'article 2 de ce travail. Ensuite, en se basant non plus sur des mesures environnementales par point mais sur des différences de valeurs environnementales entre entités spatiales, il sera possible de tester la relation entre les matrices de dissimilarité phylogénétique (remplacement spatial et gradient phylogénétique) et les matrices de différences environnementales (Melo *et al.* 2009 ; Leprieur *et al.* 2011 ; Eiserhardt *et al.* 2013). Dans la mesure où la limitation de la dispersion des organismes en fonction de la distance peut générer de l'autocorrélation spatiale, ajouter dans le modèle une matrice de distance géographique entre entités spatiales permettra d'estimer la part de la variance des composantes de la phylobétadiversité expliquée soit par des contraintes à la dispersion soit par des contraintes à l'évolution de la niche (Swenson 2011 ; Hardy *et al.* 2012 ; Eiserhardt *et al.* 2013).

5.3.3) S'affranchir des métriques de diversité afin d'évaluer l'influence de l'énergie sur le taux d'évolution moléculaire

Cette troisième perspective propose des pistes de réflexion pour mettre en relation directement les facteurs environnementaux et les processus sans recourir aux métriques de diversité, tout en conservant de larges échelles spatiales propres à la macro-écologie. Parmi les nombreuses possibilités, je propose d'évaluer l'influence de facteurs environnementaux et plus particulièrement celui de l'énergie sur le taux d'évolution moléculaire. Le taux d'évolution moléculaire correspond à la vitesse de changement de nucléotides d'une séquence d'ADN. Il dépend du produit entre la taille efficace de la population et le taux de mutation. Jusqu'à présent, les rares études ont essentiellement cherché des liens entre l'énergie ambiante (e.g. température), le taux de mutation et le processus de spéciation afin d'expliquer, par exemple, le patron latitudinal de richesse (Rohde 1992 ; Allen *et al.* 2006 ; Bromham & Cardillo, 2003). Rappelons que le taux de mutation peut être directement sous la dépendance des

rayonnements ultraviolets et surtout de la température ambiante *via* l'accélération du métabolisme chez les ectothermes et la genèse de métabolites mutagènes (radicaux libres ; Martin, 1995 ; Allen *et al.* 2002 ; Brown *et al.* 2004 ; Evans *et al.* 2005). Les liens entre température et taux de mutation peuvent être largement indirects. La diminution de taille des organismes aux températures les plus fortes (règle de Bergmann) entraine une cascade de modifications des traits biologiques (augmentation du nombre de descendants par occasion de reproduction, un voltinisme plus important et une diminution du temps de génération) qui *in fine* accélère le taux de mutation en augmentant la probabilité d'erreur de copie de l'ADN lors de la méiose (Evans *et al.* 2005 ; Gillooly *et al.* 2005). Nous avons pu voir en discussion que le support de cette hypothèse sur les effets de l'énergie ambiante restait modéré en fonction des groupes taxonomiques étudiés (Bromham & Cardillo, 2003 ; Davies *et al.* 2004 ; Lanfear *et al.* 2010 ; Goldie *et al.* 2011).

Etonnamment, très peu d'études se sont intéressées à comprendre plus généralement les effets de l'énergie productive (quantité de ressources trophiques disponibles) sur les taux d'évolution moléculaire (Davies *et al.* 2004). La quantité d'énergie productive étant positivement corrélée à la disponibilité en eau et à l'énergie ambiante, il est attendu qu'une augmentation d'énergie productive s'accompagne également d'une accélération des taux de mutation. De plus, d'abondantes quantités de ressources trophiques favorisent le maintien de taille de population importante et limitent ainsi l'extinction (Evans *et al.* 2005). La persistance d'un grand nombre d'individus soumis à des températures plus élevées pourrait jouer sur les taux d'évolution moléculaire par une augmentation conjointe de la taille de population efficace et du taux de mutation.

L'objectif de cette perspective est double, il s'agira dans un premier temps d'évaluer la corrélation entre l'énergie productive et le taux d'évolution moléculaire, puis dans un deuxième temps d'estimer la contribution relative de la taille efficace des populations et du taux de mutation dans la relation entre énergie productive et taux d'évolution moléculaire. Bien que cette perspective soit très générale et puisse être étendue à un grand nombre de groupes, une mise en pratique est illustrée à l'aide d'organismes du milieu aquatique souterrain.

Encore une fois, le genre *Proasellus* constitue un modèle biologique intéressant de par l'abondance des espèces souterraines à petites aires de répartition présentes dans une grande partie de l'Europe. En s'appuyant sur la phylogénie du genre (Morvan, 2013) et sur la base de données d'occurrences génétiques, il s'agira d'identifier des couples d'espèces sœurs ayant de petites aires de répartition géographiquement proches. Pour chaque couple, les nouvelles méthodes de séquençage à haut débit de type transcriptomique permettront de récupérer des caractères efficaces (gènes orthologues) retrouvés chez ces espèces. Les taux de substitution, les taux de substitution synonyme (dS) et le omega (rapport entre taux de substitution non synonyme et taux de substitution synonyme dN/dS) seront les variables utilisées pour estimer respectivement le taux d'évolution moléculaire, le taux de mutation et la taille efficace des populations (Kimura 1983 ; Lanfear *et al.* 2010). Afin de ne pas être influencé par la variabilité des temps de divergence entre couple d'espèces sœurs, les estimations des variables seront ramenées à la même unité temporelle à partir d'un arbre phylogénétique daté provenant d'une inférence Bayésienne (Davies *et al.* 2004 ; Morvan *et al.* 2013). Les cartes de distribution des espèces permettront de déterminer le centre de l'aire de répartition des espèces de chaque couple. Les valeurs des différentes variables moléculaires seront cartographiées afin de fournir des patrons de taux de mutation et de taille de population efficace à l'échelle de l'Europe. Les valeurs des variables environnementales notamment d'énergie productive et d'énergie ambiante seront moyennées sur les aires de répartition de chaque couple d'espèces sœurs. Enfin, sous l'hypothèse que les variations géographiques des variables environnementales sont restées stables au

cours du temps, des analyses corrélatives permettront d'établir les relations entre variables moléculaires et environnementales.

Bilbiographie

Abrams KM, Guzik MT, Cooper SJB *et al.* (2012) Molecular Phylogenetics and Evolution What lies beneath : Molecular phylogenetics and ancestral state reconstruction of the ancient subterranean Australian Parabathynellidae (Syncarida , Crustacea). *Molecular Phylogenetics and Evolution*, **64**, 130–144.

Allen AP, Brown JH, Gillooly JF (2002) Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science*, **297**, 1545–1548.

Allen AP, Gillooly JF, Savage VM, Brown JH (2006) Kinetic effects of temperature on rates of genetic divergence and speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 9130–9135.

Allicock OM, Lemey P, Tatem AJ *et al.* (2012) Phylogeography and population dynamics of dengue viruses in the Americas. *Molecular Biology and Evolution*, **29**, 1533–1543.

Anderson MJ, Crist TO, Chase JM *et al.* (2011) Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecology Letters*, **14**, 19–28.

Angilletta MJ, Niewiarowski PH, Navas CA (2002) The evolution of thermal physiology in ectotherms. *Journal of Thermal Biology*, **27**, 249–268.

Araújo MB, Nogués-Bravo D, Diniz-Filho JAF *et al.* (2008) Quaternary climate changes explain diversity among reptiles and amphibians. *Ecography*, **31**, 8–15.

Arens NC, West ID (2008) Press-pulse : a general theory of mass extinction ? *Paleobiology*, **34**, 456–471.

Arita HT, Christen JA, Rodríguez P, Soberón J (2008) Species diversity and distribution in presence-absence matrices: mathematical relationships and biological implications. *The American Naturalist*, **172**, 519–532.

Avise JC, (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge.

Avise JC, (2009) Phylogeography: retrospect and prospect. *Journal of biogeography*, **36**, 3-15.

Avise JC, Arnold J, Ball RM *et al.* (1987) Intraspecific Phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.

Barlow A, Baker K, Hendry CR *et al.* (2013) Phylogeography of the widespread African puff adder (Bitis arietans) reveals multiple Pleistocene refugia in southern Africa. *Molecular Ecology*, **22**, 1134–1157.

Barr TCJ, Holsinger JR (1985) Speciation in cave faunas. *Annual Review of Ecology and Systematics*, **16**, 313–337.

Baselga A (2010a) Multiplicative partition of true diversity yields independent alpha and beta components ; additive partition does not. *Ecology*, **91**, 1974–1981.

Baselga A (2010b) Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, **19**, 134–143.

Baselga A (2012) The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. *Global Ecology and Biogeography*, **21**, 1223–1232.

Baselga A, Lobo JM, Svenning J, Arau MB (2012) Global patterns in the shape of species geographical ranges reveal range determinants. *Journal of Biogeography*, **39**, 760–771.

Beaumont M a, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Beaumont MA, Nielsen R, Robert C *et al.* (2010) In defence of model-based inference in phylogeography. *Molecular Ecology*, **19**, 436–446.

Beck J, Ballesteros-Mejia L, Buchmann CM *et al.* (2012) What's on the horizon for macroecology? *Ecography*, **35**, 673–683.

Beheregaray LB, Caccone A (2007) Cryptic biodiversity in a changing world. *Journal of Biology*, **6**, 9.

Bell RC, MacKenzie JB, Hickerson MJ *et al.* (2012) Comparative multi-locus phylogeography confirms multiple vicariance events in co-distributed rainforest frogs. *Proceedings. Biological sciences / The Royal Society*, **279**, 991–999.

Bickford D, Lohman DJ, Sodhi NS *et al.* (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, **22**, 148–155.

Blanquart F, Kaltz O, Nuismer SL, Gandon S (2013) A practical guide to measuring local adaptation. *Ecology Letters*, **16**, 1195–205.

Bloomquist EW, Lemey P, Suchard MA (2010) Three Roads Diverged? Routes To Phylogeographic Inference. *Trends in Ecology & Evolution*, **25**, 626–632.

Boakes EH, McGowan PJK, Fuller RA *et al.* (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS biology*, **8**, e1000385.

Bonn A, Storch D, Gaston KJ (2004) Structure of the species-energy relationship. *Proceedings. Biological sciences / The Royal Society*, **271**, 1685–1691.

Borregaard MK, Rahbek C (2010) Dispersion fields, diversity fields and null models: uniting range sizes and species richness. *Ecography*, **33**, 402–407.

Botosaneanu L (1986) *Stygofauna Mundi*. Brill EJ, Bachuys W, Leiden, Netherlands.

Boucher FC, Thuiller W, Roquet C *et al.* (2011) Reconstructing the origins of high-alpine niches and cushion life form in the genus *Androsace* s.l. (primulacea). *Evolution*, **66**, 1255–1268.

Bouckaert R, Lemey P, Dunn M *et al.* (2012) Mapping the origins and expansion of the Indo-European language family. *Science*, **337**, 957–960.

Boutin C (1994) Phylogeny and biogeography of metacrangonyctid amphipods in North Africa. *Hydrobiologia*, **287**, 49–64.

Brancelj A, Boonyanusith C, Watiroyram S, Sanoamuang L (2013) The groundwater-dwelling fauna of Southeast Asia. *Journal of Limnology*, **72**, 327–344.

Brayard A, Escarguel G, Bucher H (2005) Latitudinal gradient of taxonomic richness : combined outcome of temperature and geographic mid-domains effects ? *Journal of Zoological Systematics and Evolutionary Research*, **43**, 178–188.

Bromham L, Cardillo M (2003) Testing the link between the latitudinal gradient in species richness and rates of molecular evolution Higher speciation rate. *Journal of Evolutionary Biology*, **16**, 200–207.

Brown JH (1995) *Macroecology*. University of Chicago Press.

Brown DM, Brenneman R a, Koepfli K-P *et al.* (2007) Extensive population genetic structure in the giraffe. *BMC Biology*, **5**, 57.

Brown JH, Gillooly JF, Allen AP, Savage VM, West GB (2004) Toward a Metabolic theory of ecology. *Ecology*, **85**, 1771–1789.

Brown JH, Maurer BA (2007) Macroecology: the division of food and space among species on continents. *Science*, **243**, 1145–1150.

Brunsdon C, Fotheringham AS, Charlton ME (1996) Geographically Weighted Regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, **28**, 281–298.

Bryant JA, Lamanna C, Morlon H *et al.* (2008) Microbes on mountainsides : Contrasting elevational patterns of bacterial and plant diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 11505–11511.

Buckley LB, Jetz W (2007) Environmental and historical constraints on global patterns of amphibian richness. *Proceedings. Biological sciences / The Royal Society*, **274**, 1167–1173.

Cadotte MW, Jonathan Davies T, Regetz J *et al.* (2010) Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history. *Ecology Letters*, **13**, 96–105.

Carnaval AC, Hickerson MJ, Haddad CFB (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Nature*, **323**, 785–789.

Carstens B, Lemmon AR, Lemmon EM (2012) The promises and pitfalls of next-generation sequencing data in phylogeography. *Systematic Biology*, **61**, 713–715.

Carstens BC, Richards CL (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**, 1439–1454.

Carvalho JC, Cardoso P, Gomes P (2012) Determining the relative roles of species replacement and species richness differences in generating beta-diversity patterns. *Global Ecology and Biogeography*, **21**, 760–771.

Castany G (1998) *Hydrogéologie : principes et méthodes*. Dunod, Paris.

Castellarini F, Dole-Olivier M-J, Malard F, Gibert J (2007a) Using habitat heterogeneity to assess stygobiotic species richness in the French Jura region with a conservation perspective. *Fundamental and Applied Limnology*, **169**, 69–78.

Castellarini F, Malard F, Dole-Olivier M-J, Gibert J (2007b) Modelling the distribution of stygobionts in the Jura Mountains (eastern France). Implications for the protection of ground waters. *Diversity and Distributions*, **13**, 213–224.

Cavender-Bares J, Ackerly DD, Baum DA, Bazzaz FA (2004) Phylogenetic overdispersion in Floridian oak communities. *The American Naturalist*, **163**, 823–843.

Cavender-Bares J, Gonzalez-Rodriguez A, Pahlich A, Koehler K, Deacon N (2011) Phylogeography and climatic niche evolution in live oaks (*Quercus* series *Virentes*) from the tropics to the temperate zone. *Journal of Biogeography*, **38**, 962–981.

Cavender-Bares J, Keen A, Miles B (2006) Phylogenetic structure of Floridian plant communities depends on Taxonomic and spatial scale. *Ecology*, **87**, S109–S122.

Cavender-Bares J, Kozak KH, Fine PV a, Kembel SW (2009) The merging of community ecology and phylogenetic biology. *Ecology Letters*, **12**, 693–715.

CDB (2010) Objectifs d'Aichi pour la biodiversité programme stratégique pour la biodiversité 2011-2020. Convention sur la diversité biologique. http://www.cbd.int/sp/targets/ .

Chan LM, Brown JL, Yoder AD (2011) Integrating statistical genetic and geospatial methods brings new power to phylogeography. *Molecular Phylogenetics and Evolution*, **59**, 523–537.

Chao A, Gotelli NJ, Hsieh TC *et al.* (2014) Rarefaction and extrapolation with Hill numbers : a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, **84**, 45–67.

Chao A, Jost L (2012) Coverage-based rarefaction and extrapolation : standardizing samples by completeness rather than size. *Ecology*, **93**, 2533–2547.

Chave J (2013) The problem of pattern and scale in ecology: what have we learned in 20 years? *Ecology Letters*, **16**, 4–16.

Christman MC, Culver DC (2001) The relationship between cave biodiversity and available habitat. *Journal of Biogegraphy*, **28**, 367–380.

Christman MC, Culver DC, Madden MK, White D (2005) Patterns of endemism of the eastern North American cave fauna. *Journal of Biogeography*, **32**, 1441–1452.

Clark JS (2009) Beyond neutral science. *Trends in Ecology & Evolution*, **24**, 8–15.

Coineau N (1994) Evolutionary biogeography of the microparasellid isopod *Microcharon* (Crustacea) in the Mediterranean Basin. *Hydrobiologia*, **287**, 77–94.

Colwell RK, Coddington J (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **345**, 101–118.

Colwell RK, Gotelli NJ, Rahbek C *et al.* (2009) Peaks, plateaus, canyons, and craters : the complex geometry of simple mid-domain effect models. *Evolutionary Ecology Research*, **11**, 355–370.

Colwell RK, Lees DC (2000) The mid-domain effect : geometric species richness. *Trends in Ecology & Evolution*, **15**, 70–76.

Colwell RK, Rahbek C, Gotelli NJ (2004) The Mid-Domain Effect and Species Richness Patterns : What Have We Learned So Far ? *The American Naturalist*, **163**, E1–E23.

Commision for the Geological Map of the World. 1999. Maps of the world environment during the last two climatic extremes. The last glacial maximum. Geoportal http://www.ukge.com/en-gb/World-Environments-during-the-Last-Two-Climatic-Extremes__p-1812.aspx .

Condamine FL, Rolland J, Morlon H (2013) Macroevolutionary perspectives to environmental change. *Ecology Letters*, **16**, 72–85.

Condamine FL, Sperling FAH, Wahlberg N, Rasplus J-Y, Kergoat GJ (2012) What causes latitudinal gradients in species diversity? Evolutionary processes and ecological constraints on swallowtail biodiversity. *Ecology Letters*, **15**, 267–277.

Cordellier M, Pfenninger M (2009) Inferring the past to predict the future: climate modelling predictions and phylogeography for the freshwater gastropod Radix balthica (Pulmonata, Basommatophora). *Molecular Ecology*, **18**, 534–544.

Costa GC, Wolfe C, Shepard DB, Caldwell JP, Vitt LJ (2008) Detecting the influence of climatic variables on species distributions: a test using GIS niche-based models along a steep longitudinal environmental gradient. *Journal of Biogeography*, **35**, 637–646.

Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410–418.

Culver DC (2005) Life history, evolution. Dans : *Encyclopedia of caves* (eds Culver DC & White WB), pp 346-349. Elsevier, London.

Culver DC, Christman MC, Sket B, Trontelj P (2004) Sampling adequacy in an extreme environment : species richness patterns in Slovenian caves. *Biodiversity and Conservation*, **13**, 1209–1229.

Culver DC, Deharveng L, Bedos A *et al.* (2006) The mid-latitude biodiversity ridge in terrestrial cave fauna. *Ecography*, **29**, 120–128.

Culver DC, Pipan T, Schneider K (2009) Vicariance, dispersal and scale in the aquatic subterranean fauna of karst regions. *Freshwater Biology*, **54**, 918–929.

Culver DC, Sket B (2000) Hotspots of subterranean biodiversity in caves and wells. *Journal of Cave and Karst Studies*, **62**, 11–17.

Currie DJ, Mittelbach GG, Cornell H V. *et al.* (2004) Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters*, **7**, 1121–1134.

Danielopol D, Griebler C, Gunatilaka A, Notenboom Jost (2003) Present state and future prospects for groundwater ecosystems. *Environmental Conservation*, **30**, 1–27.

Darwin C. 1862. *The Voyage of the Beagle*. Garden City, NJ: Doubleday

Datry T, Malard F, Gibert J (2005) Response of invertebrate assemblages to increased groundwater recharge rates in a phreatic aquifer. *Journal of the North American Benthological Society*, **24**, 461–477.

Davies JT (2006) Evolutionary ecology: when relatives cannot live together. *Current Biology*, **16**, R643–R647.

Davies R, Orme C, Storch D *et al.* (2007a) Topography, energy and the global distribution of bird species richness. *Proceedings of the Royal Society B*, **274**, 1189–1197.

Davies RG, Orme CDL, Webster AJ *et al.* (2007b) Environmental predictors of global parrot (Aves: Psittaciformes) species richness and phylogenetic diversity. *Global Ecology and Biogeography*, **16**, 220–233.

Davies TJ, Buckley LB (2011) Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present-day richness gradients for mammals. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, **366**, 2414–2425.

Davies TJ, Buckley LB, Grenyer R, Gittleman JL (2011) The influence of past and present climate on the biogeography of modern mammal diversity. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, **366**, 2526–2535.

Davies TJ, Savolainen V, Chase MW, Moat J, Barraclough TG (2004) Environmental energy and evolutionary rates in flowering plants. *Proceedings. Biological sciences / The Royal Society*, **271**, 2195–2200.

Deharveng L, Stoch F, Gibert J *et al.* (2009) Groundwater biodiversity in Europe. *Freshwater Biology*, **54**, 709–726.

Deharveng L (2005) Diversity patterns in the tropics. Dans *Encyclopedia of caves* (Culver DC & White WB), pp 166-170, Elsevier, London.

Dias MS, Cornu J-F, Oberdorff T, Lasso CA, Tedesco PA (2013) Natural fragmentation in river networks as a driver of speciation for freshwater fishes. *Ecography*, **36**, 683–689.

Diniz-Filho JAF, Loyola RD, Raia P, Mooers AO, Bini LM (2013) Darwinian shortfalls in biodiversity conservation. *Trends in Ecology & Evolution*, **28**, 689–695.

Dole-Olivier M-J, Castellarini F, Coineau N *et al.* (2009a) Towards an optimal sampling strategy to assess groundwater biodiversity: comparison across six European regions. *Freshwater Biology*, **54**, 777–796.

Dole-Olivier M-J, Malard F, Martin D, Lefébure T, Gibert J (2009b) Relationships between environmental variables and groundwater biodiversity at the regional scale. *Freshwater Biology*, **54**, 797–813.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.

Dunn RR, McCain CM, Sanders NJ (2007) When does diversity fit null model predictions? Scale and range size mediate the mid-domain effect. *Global Ecology and Biogeography*, **16**, 305–312.

Dynesius M, Jansson R (2000) Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 9115–9120.

Edwards CJ, Suchard MA, Lemey P *et al.* (2011) Article ancient hybridization and an Irish origin for the modern polar bear matriline. *Current Biology*, **21**, 1251–1258.

Ehlers J, Gibbard PL (2004) *Quaternary glaciations - extent and chronology.* Part I: Europe. Development in Quaternary Science 2. Elsevier.

Eiserhardt WL, Bjorholm S, Svenning J-C, Rangel TF, Balslev H (2011) Testing the water-energy theory on American palms (Arecaceae) using geographically weighted regression. *PloS one*, **6**, e27027.

Eiserhardt WL, Svenning J-C, Baker WJ, Couvreur TLP, Balslev H (2013) Dispersal and niche evolution jointly shape the geographic turnover of phylogenetic clades across continents. *Scientific reports*, **3**, 1164.

Emerson BC, Cicconardi F, Fanciulli PP, Shaw PJA (2011) Phylogeny, phylogeography, phylobetadiversity and the molecular analysis of biological communities. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, **366**, 2391–2402.

Engel AS (2005) Chemoautotrophy. Dans *Encyclopedia of caves* (Culver DC & White WB), pp 90-102, Elsevier, London.

Erwin DH (2000) Macroevolution is more than repeated rounds of microevolution. *Evolution & Developpment*, **84**, 78–84.

Escarguel G, Brayard a., Bucher H (2008) Evolutionary rates do not drive latitudinal diversity gradients. *Journal of Zoological Systematics and Evolutionary Research*, **46**, 82–86.

Evans KL, Warren PH, Gaston KJ (2005) Species–energy relationships at the macroecological scale: a review of the mechanisms. *Biological Reviews*, **80**, 1–25.

Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology*, **13**, 853–864.

Excoffier L, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 481–501.

Fagundes NJR, Ray N, Beaumont M *et al.* (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 17614–17619.

Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation*, **61**, 1–10.

Faria NR, Suchard M a, Abecasis A *et al.* (2012) Phylodynamics of the HIV-1 CRF02_AG clade in Cameroon. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, **12**, 453–460.

Felsenstein J (1985) Phylogenies and the comparative method. The *American Naturalist*, **125**, 1–15.

Ferreira D (2005) *Biodiversité aquatique souterraine de France : base de données, patrons de distribution et implications en termes de conservation*. Thèse de doctorat, Université Claude Bernard Lyon1.

Ferreira D, Malard F, Dole-Olivier M, Gibert J (2007) Obligate groundwater fauna of France : diversity patterns and conservation implications. *Biodiversity and Conservation*, **16**, 567–596.

Field R, Hawkins BA, Cornell H V. *et al.* (2009) Spatial species-richness gradients across scales: a meta-analysis. *Journal of Biogeography*, **36**, 132–147.

Finston T, Johnson M, Humphreys W, Eberhard S, Halse S (2007) Cryptic speciation in two widespread subterranean amphipod genera reflects historical drainage patterns in an ancient landscape. *Molecular Ecology*, **16**, 355–365.

Foody GM (2004) Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Global Ecology and Biogeography*, **13**, 315–320.

Forest F, Grenyer R, Rouget M *et al.* (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature*, **445**, 757–760.

Fotheringham AS, Brunsdon C, Charlton M (2002) *Geographically weighted regression: the analysis of spatially varying relationships*. Wiley, Chichester.

Foulquier A, Malard F, Mermillod-Blondin F *et al.* (2010) Vertical change in dissolved organic carbon and oxygen at the water table region of an aquifer recharged with stormwater: biological uptake or mixing? *Biogeochemistry*, **99**, 31–47.

Foulquier A, Mermillod-Blondin F, Malard F, Gibert J (2011) Response of sediment biofilm to increased dissolved organic carbon supply in groundwater artificially recharged with stormwater. *Journal of Soils and Sediments*, **11**, 382–393.

Freeze RA, Cherry JA (1979) *Groundwater*. Prentice-Hall, Englewood Cliffs.

Frichot E, Schoville SD, Bouchard G, Fran O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30**, 1687–1699.

Gaston KJ (2000) Global patterns in biodiversity. *Nature*, **405**, 220–227.

Gaston KJ, Blackburn TM, Spicer JI (1998) Rapoport's rule : time for an epitaph ? *Trends in Ecology & Evolution*, **13**, 70–74.

Gaston KJ, Chown SL, Calosi P *et al.* (2009) Macrophysiology: a conceptual reunification. *The American Naturalist*, **174**, 595–612.

Gaston KJ, Fuller R a. (2009) The sizes of species' geographic ranges. *Journal of Applied Ecology*, **46**, 1–9.

Ghalambor CK, Huey RB, Martin PR, Tewksbury JJ, Wang G (2006) Are mountain passes higher in the tropics? Janzen's hypothesis revisited. *Integrative and Comparative Biology*, **46**, 5–17.

Gibert J, Brancelj A, Camacho A, *et al.* (2005) Groundwater Biodiversity. Protocols for the ASsessment and Conservation of Aquatic Life In the Subsurface (PASCALIS): overview and main results. Dans J. Gibert, éditeur, *Proceedings on symposium on World Subterranean Biodiversity* (eds Gibert J), Université Claude Bernard Lyon1, France.

Gibert J, Culver DC (2005) Diversity patterns in Europe. Dans *Encyclopedia of caves* (eds Culver DC & White WB), pp 196-201. Elsevier, London.

Gibert J, Culver DC, Dole-Olivier M-J *et al.* (2009) Assessing and conserving groundwater biodiversity: synthesis and perspectives. *Freshwater Biology*, **54**, 930–941.

Gibert J, Deharveng L (2002) Subterranean Ecosystems: A Truncated Functional Biodiversity. *BioScience*, **52**, 473–481.

Gill MS, Lemey P, Faria NR *et al.* (2013) Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Molecular Biology and Evolution*, **30**, 713–724.

Gillooly JF, Allen AP, West GB, Brown JH (2005) The rate of DNA evolution : Effects of body size and temperature on the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 140–145.

Gillooly JF, Charnov EL, West GB, Savage VM, Brown JH (2002) Effects of size and temperature on developmental time. *Nature*, **417**, 70–73.

Gilpin E, Soulé ME (1986) Minimum Viable Populations: processus d'extinction des espèces. Dans *Conservation Biology: The Science of Scarcity and Diversity* (eds Soulé ME), pp 19-34, Sinauer, Sunderland, Mass.

Ginet R, Decou V (1977) *Initiation à la biologie et à l'écologie souterraines*. Delarge, Paris.

GlobCover, (2009) European Space Agengy ESA, Portal http://due.esrin.esa.int/globcover/

Gould SC, Lewontin RC (1979) The spandrels of san marco and the panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, **205**, 581–598.

Goldie X, Lanfear R, Bromham L (2011) Diversification and the rate of molecular evolution: no evidence of a link in mammals. *BMC Evolutionary Biology*, **11**, 286.

Gotelli NJ, Anderson MJ, Arita HT *et al.* (2009) Patterns and causes of species richness: a general simulation model for macroecology. *Ecology Letters*, **12**, 873–886.

Gotelli NJ, Colwell RK (2001) Quantifying biodiversity : procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.

Gouveia SF, Hortal J, Cassemiro FAS, Rangel TF, Diniz-Filho JAF (2013) Nonstationary effects of productivity, seasonality, and historical climate changes on global amphibian diversity. *Ecography*, **36**, 104–113.

Graham CH, Fine PVA (2008) Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecology Letters*, **11**, 1265–1277.

Grinnell J (1917) The niche-relationships of the california thrasher. *Auk*, **34**, 427–433.

Guisan A, Edwards TC (2002) Generalized linear and generalized additi v e models in studies of species distributions : setting the scene. *Ecological Modelling*, **157**, 89–100.

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Hardy OJ, Couteron P, Munoz F, Ramesh BR, Pélissier R (2012) Phylogenetic turnover in tropical tree communities: impact of environmental filtering, biogeography and mesoclimatic niche conservatism. *Global Ecology and Biogeography*, **21**, 1007–1016.

Hardy OJ, Senterre B (2007) Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. *Journal of Ecology*, **95**, 493–506.

Hawkins BA, Albuquerque FS, Araújo MB *et al.* (2007) A global evaluation of metabolic theory as an explanation for terrestrial species richness gradients. *Ecology*, **88**, 1877–1888.

Hawkins BA, Field R, Cornell H V. *et al.* (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105–3117.

Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, **8**, 289.

Helmus MR, Bland TJ, Williams CK, Ives AR (2007) Phylogenetic Measures of Biodiversity. *The American Naturalist*, **169**.

Henry JP (1976) *Recherches sur les Asellidae hypogées de la lignée cavaticus (Crustacea, Isopoda, Asellota)*. Thèse de doctorat, Université de Dijon.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hey J (2006) On the failure of modern species concepts. *Trends in Ecology & Evolution*, **21**, 447–450.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–920.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2785–2790.

Hickerson MJ, Carstens BC, Cavender-Bares J *et al.* (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution*, **54**, 291–301.

Hickerson MJ, Meyer CP (2008) Testing comparative phylogeographic models of marine vicariance and dispersal using a hierarchical Bayesian approach. *BMC Evolutionary Biology*, **8**, 322.

Hickerson MJ, Stahl EA, A. LH (2006) Test for simultaneous divergence using approximate Bayesian computation. *Evolution*, **60**, 2435–2453.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hillebrand H (2004) On the Generality of the Latitudinal Diversity Gradient. *The American Naturalist*, **163**, 192–211.

Ho SYW, Saarma U, Barnett R, Haile J, Shapiro B (2008) The effect of inappropriate calibration: three case studies in molecular ecology. *PloS one*, **3**, e1615.

Hof C, Brändle M, Brandl R (2008) Latitudinal variation of diversity in European freshwater animals is not concordant across habitat types. *Global Ecology and Biogeography*, **17**, 539–546.

Holsinger JR (2005) Vicariance and dispersalist biogeography. Dans *Encyclopedia of caves* (eds Culver DC & White WB), pp 591-599. Elsevier, London.

Hortal J, Diniz-Filho JAF, Bini LM *et al.* (2011) Ice age climate, evolutionary constraints and diversity patterns of European dung beetles. *Ecology Letters*, **14**, 741–748.

Howarth FG (1987) The evolution of non-relictual tropical troglobites. *International Journal of Speleology*, **16**, 1–16.

Hubbell SP (2001). *A unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ.

Hughes PD, Woodward JC (2008) Timing of glaciation in the Mediterranean mountains during the last cold stage. *Journal of Quaternary Science* **23**, 575-588.

Hugueny B, Movellan A, Belliard J (2011) Habitat fragmentation and extinction rates within freshwater fish communities: a faunal relaxation approach. *Global Ecology and Biogeography*, **20**, 449–463.

Hüppop K (2005) Adaptation to low food. Dans *Encyclopedia of caves* (eds Culver DC & White WB), pp 4-10. Elsevier, London.

Hutchinson GE (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.

Illies J (1978) *Limnofauna Europaea*. 2nd edn. Gustav Fischer Verlag, Stuttgart.

Jablonski D, Roy K, Valentine JW (2006) Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. *Science*, **314**, 102–106.

Jansson R, Dynesius M (2002) The fate of clades in a world of recurrent climatic change: Milankovitch oscillations and evolution. *Annual Review of Ecology and Systematics*, **33**, 741–777.

Janzen DH (1967) Why mountain pass are higher in the tropics. *The American naturalist*, **101**, 233–249.

Jenkins DG, Ricklefs RE (2011) Biogeography and ecology: two views of one world. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **366**, 2331–2335.

Jetz W, McPherson JM, Guralnick RP (2012) Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in Ecology & Evolution*, **27**, 151–159.

Jetz W, Rahbek C (2002) Geographic Range Size and Determinants of Avian Species Richness. *Science*, **297**, 1548–1551.

Jocque M, Field R, Brendonck L, De Meester L (2010) Climatic control of dispersal-ecological specialization trade-offs: a metacommunity process at the heart of the latitudinal diversity gradient? *Global Ecology and Biogeography*, **19**, 244–252.

Jost L (2005) Partitioning diversity into independent alpha and beta components. *Ecology*, **88**, 2427–2439.

Juberthie C, Decu V (1994) *Encyclopedia Biospeleologica I*, volume 1. Société de biospéléologie, Moulis.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, 7, 1225–1241.

Kearney M (2006) Habitat, environment and niche: what are we modelling ? *Oikos*, **115**, 186–191.

Kearney M, Porter W (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, **12**, 334–350.

Keith SA, Connolly SR (2013) Effects of diversity-dependent colonization-extinction dynamics on the mid-domain effect. *Global Ecology and Biogeography*, **22**, 773–783.

Keppel G, Van Niel KP, Wardell-Johnson GW *et al.* (2012) Refugia: identifying and understanding safe havens for biodiversity under climate change. *Global Ecology and Biogeography*, **21**, 393–404.

Kerr JT, Packer L (1997) Habitat heterogeneity as a determinant of mammal species richness in high-energy regions. *Nature*, **385**, 252–254.

Kerr JT, Southwood TR, Cihlar J (2001) Remotely sensed habitat diversity predicts butterfly species richness and community similarity in Canada. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 11365–11370.

Kidd DM, Ritchie MG (2006) Phylogeographic information systems: putting the geography into phylogeography. *Journal of Biogeography*, **33**, 1851–1865.

Kimura M (1983) *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.

Kingman JFC (1982a) The coalescent. *Stochastic Processes and their Applications*, **13**, 235–248.

Kingman JFC (1982b) On the genealogy of large populations. *Journal of Applied Probability*, **19**, 27–43.

Koleff P, Gaston KJ, Lennon JJ (2003) Measuring beta diversity for presence – absence data. *Journal of Animal Ecology*, **72**, 367–382.

Kuhner MK (2009) Coalescent genealogy samplers: windows into population history. *Trends in Ecology & Evolution*, **24**, 86–93.

Landis MJ, Matzke NJ, Moore BR, Huelsenbeck JP (2013) Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology*, **62**, 789-804.

Lanfear R, Ho SYW, Love D, Bromham L (2010) Mutation rate is linked to diversification in birds. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 20423–20428.

Lauzerale C (2012) *Prédiction du potentiel d'invasion des espèces non natives par des modèles de niche : approches méthodologiques et applications aux poissons d'eau douce sur le territoire français*. Thèse de Doctorat, Université de Toulouse III Paul Sabatier.

Lavergne S, Mouquet N, Thuiller W, Ronce O (2010) Biodiversity and Climate Change: Integrating Evolutionary and Ecological Responses of Species and Communities. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 321–350.

Lawton JH (1999) Are there general laws in ecology ? *Oikos*, **84**, 177–192.

Lefébure T (2005) *Origine, évolution et mesure de la biodiversité des eaux souterraines : analyse moléculaire du genre Niphargus (Crustacea)*. Thèse de doctorat, Université Claude Bernard Lyon 1.

Lefébure T, Douady CJ, Gouy M *et al.* (2006a) Phylogeography of a subterranean amphipod reveals cryptic diversity and dynamic evolution in extreme environments. *Molecular Ecology*, **15**, 1797–1806.

Lefébure T, Douady CJ, Gouy M, Gibert J (2006b) Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution*, **40**, 435–447.

Lefébure T, Douady CJ, Malard F, Gibert J (2007) Testing dispersal and cryptic diversity in a widely distributed groundwater amphipod (Niphargus rhenorhodanensis). *Molecular Phylogenetics and Evolution*, **42**, 676–686.

Leibold M a., Holyoak M, Mouquet N *et al.* (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, **7**, 601–613.

Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. *Plos Computational Biology*, **5**, e1000520.

Lemey P, Rambaut A, Welch JJ, Suchard MA (2010) Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution*, **27**, 1877–1885.

Lemmon AR, Lemmon EM (2008) A likelihood framework for estimating phylogeographic history on a continuous landscape. *Systematic biology*, **57**, 544–651.

Lennon JJ, Koleff P, Greenwood JJD, Gaston KJ (2003) Contribution of rarity and commonness to patterns of species richness. *Ecology Letters*, **7**, 81–87.

Leprieur F, Albouy C, De Bortoli J *et al.* (2012) Quantifying phylogenetic beta diversity: distinguishing between "true" turnover of lineages and phylogenetic diversity gradients. *PloS one*, **7**, e42760.

Leprieur F, Tedesco PA, Hugueny B *et al.* (2011) Partitioning global patterns of freshwater fish beta diversity reveals contrasting signatures of past climate changes. *Ecology Letters*, **14**, 325–334.

Lewin R. (1989) Biologists disagree over bold signature of nature. *Science*, **5**, 527-528

Lindgren DAG (1972) The temperature influence on the spontaneous mutation rate. *Heritas*, **178**, 165–172.

Lorenzen ED, Nogués-Bravo D, Orlando L *et al.* (2011) Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*, **479**, 359 – 364.

Losos JB, Schluter D (2000) Analysis of an evolutionary species-area relationship. *Nature*, **408**, 847–850.

Lozupone C, Knight R (2005) UniFrac : a new phylogenetic method for comparing microbial communities. *Applied and Environnemental Microbiology*, **71**, 8228–8235.

MacArthur RH, Wilson EO (1967) *The theory of island biogeography*. Princeton University Press.

Magurran AE (2004) *Measuring biological diversity*. Blackwell, Oxford, UK.

Malard F, Boutin C, Camacho A *et al.* (2009) Diversity patterns of stygobiotic crustaceans across multiple spatial scales in Europe. *Freshwater Biology*, **54**, 756–776.

Malard F, Cornu JF, Eme D, Fiser C, Galassi D, Marmonier P, Stoch F, Zagmajster M (2012) Filling the gap: distribution database of European groundwater crustaceans. *BioFresh Newsletter* 4:7. http://www.freshwaterbiodiversity.eu/tl_files/downloads/Downloads/ BioFresh_newsletter_issue_004.pdf , Accession mars 2013.

Malard F, Hervant F (1999) Oxygen supply and the adaptations of animals in groundwater. *Freshwater Biology*, **41**, 1–30.

Marmonier P, Vervier P, Gibert J, Dole-Olivier M-J (1993) Biodiversity in groundwaters. *Trends in Ecology & Evolution* , **8**,392–395.

Marske KA, Rahbek C, Nogués-Bravo D (2013) Phylogeography: spanning the ecology-evolution continuum. *Ecography*, **36**, 1169–1181.

Martin AP (1995) Metabolic rate and directional nucleotide substitution in animal Mitochondrial DNA. *Molecular Biology and Evolution*, **12**, 1124–1131.

Martin P, De Broyer C, Fiers F *et al.* (2009) Biodiversity of Belgian groundwater fauna in relation to environmental conditions. *Freshwater Biology*, **54**, 814–829.

Martins, EP, Hansen, TF (1997) Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of inter-specific data. *The American Naturalist*, **149**, 646–667.

McCain CM (2009) Vertebrate range sizes indicate that mountains may be “higher” in the tropics. *Ecology letters*, **12**, 550–560.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.

McGill BJ, Maurer BA, Weiser M (2006) Empirical evaluation of neutral theory. *Ecology*, **87**, 1411–1423.

McInerney CE, Maurice L, Robertson AL *et al.* (2014) The ancient Britons: groundwater fauna survived extreme climate change over tens of millions of years across NW Europe. *Molecular Ecology*, **23**, 1153–1166.

Melo AS, Rangel TFLVB, Diniz-Filho JAF (2009) Environmental drivers of beta-diversity patterns in New-World birds and mammals. *Ecography*, **32**, 226–236.

Mermillod-Blondin F, Lefour C, Lalouette L *et al.* (2013) Thermal tolerance breadths among groundwater crustaceans living in a thermally constant environment. *The Journal of Experimental Biology*, **216**, 1683–1694.

Michel G, Malard F, Deharveng L *et al.* (2009) Reserve selection for conserving groundwater biodiversity. *Freshwater Biology*, **54**, 861–876.

Minin VN, Bloomquist EW, Suchard MA (2008) Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, **25**, 1459–1471.

Mittelbach GG, Schemske DW, Cornell H V *et al.* (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, **10**, 315–331.

Monaghan MT, Wild R, Elliot M *et al.* (2009) Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic Biology*, **58**, 298–311.

Morin X, Thuiller W (2009) Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change. *Ecology*, **90**, 1301–1313.

Moritz C, Langham G, Kearney M *et al.* (2012) Integrating phylogeography and physiology reveals divergence of thermal traits between central and peripheral lineages of tropical rainforest lizards. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, **367**, 1680–1687.

Morlon H (2014) Phylogenetic approaches for studying diversification. *Ecology Letters*, **17**, 508–525.

Morlon H, Potts MD, Plotkin JB (2010) Inferring the dynamics of diversification: a coalescent approach. *PLoS biology*, **8**, e1000493.

Morlon H, Schwilk DW, Bryant J a *et al.* (2011) Spatial patterns of phylogenetic diversity. *Ecology Letters*, **14**, 141–149.

Morris BL, Lawrence ARL, Chilton PJC, Adams B, Calow RC, Klinck BA (2003) Groundwater and its susceptibility to degradation: a global assessment of the problem and options for management. *Early warning and assessment report series, RS. 03 – 3. UNEP*, Nairobi, Kenya

Morrone JJ, Crisci J V (1995) Historical biogeography : introduction to methods. *Annual Review of Ecology and Systematics*, **26**, 373–401.

Morueta-Holme N, Enquist BJ, McGill BJ *et al.* (2013) Habitat area and climate stability determine geographical variation in plant species range sizes. *Ecology Letters*, **16**, 1446–1454.

Morvan C (2013) *Approche phylogénétique pour l'étude des patrons macroévolutifs et processus sous-jacents dans les eaux souterraines*. Thèse de doctorat, Université Claude Bernard Lyon 1.

Morvan CLM, Malard FLM, Paradis EMP *et al.* (2013) Timetree of Aselloidea Reveals Species Diversification Dynamics in Groundwater. *Systematic Biology*, **62**, 512-522.

Mouchet MA, Villéger S, Mason NWH, Mouillot D (2010) Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. *Functional Ecology*, **24**, 867–876.

Mouquet N, Devictor V, Meynard CN *et al.* (2012) Ecophylogenetics: advances and perspectives. *Biological reviews of the Cambridge Philosophical Society*, **87**, 769–785.

Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. *Molecular Ecology*, **18**, 1034–1047.

Niemiller ML, Zigler KS (2013) Patterns of cave biodiversity and endemism in the Appalachians and Interior Plateau of Tennessee, USA. *PloS one*, **8**, e64177.

Nogués-Bravo D (2009) Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography*, **18**, 521–531.

O'Brien EM (2006) Biological relativity to water-energy dynamics. *Journal of Biogeography*, **33**, 1868–1888.

Oberdorff T, Tedesco PA, Hugueny B *et al.* (2011) Global and regional patterns in riverine fish species richness: a review. *International Journal of Ecology*, **2011**, 1–12.

Ohlemüller R, Anderson BJ, Araújo MB *et al.* (2008) The coincidence of climatic and species rarity: high risk to small-range species from climate change. *Biology Letters*, **4**, 568–572.

Oleksyk TK, Smith MW, O'Brien SJ (2010) Genome-wide scans for footprints of natural selection. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**, 185–205.

Orme CDL, Davies RG, Olson VA *et al.* (2006) Global patterns of geographic range size in birds. *PLoS biology*, **4**, e208.

Palmer MW (1994) Variation in species rochness: Towards a unification of hypotheses. *Folia Geobotanica & Phytotaxonomica*, **29**, 511–530.

Pavoine S, Ollier S, Dufour A-B (2005) Is the originality of a species measurable? *Ecology Letters*, **8**, 579–586.

Pearson RG, Dawson TP (2003) Predicting the impacts of climate change on the distribution of species : are bioclimate envelope models useful ? *Global Ecology and Biogeography*, **12**, 361–371.

Penny D (2005) Relativity for molecular clocks. *Nature*, **436**, 183–184.

Petermann JS, Fergus AJF, Chirstiane R *et al.* (2010) Biology, chance, or history ? The predictable reassembly of temperate grassland communities. *Ecology*, **91**, 408–421.

Petit C, Campy M, Chaline J, Bonvalot J (1996) Major palaeohydrographic changes in Alpine foreland during the Pliocene-Pleistocene. *Boreas*, **25**, 131-143.

Pfenninger M, Schwenk K (2007) Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology*, **7**, 121.

Pianka ER (1966) Latitudinal gradients in species diversity: a review of concepts. *The American Naturalist*, **100**, 33–46.

Pigot AL, Tobias JA (2013) Species interactions constrain geographic range expansion over evolutionary time. *Ecology letters*, **16**, 330–338.

Podani J, Schmera D (2011) A new conceptual and methodological framework for exploring and explaining pattern in presence - absence data. *Oikos*, **120**, 1625–1638.

Pons J, Barraclough TG, Gomez-Zurita J *et al.* (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–609.

Porter ML (2007) Subterranean biogeography: what have we learned from molecular techniques? *Journal of Cave and Karst Studies*, **69**, 179–186.

Poulin R (1995) Phylogeny , Ecology , and the Richness of Parasite Communities in Vertebrates. *Ecological Monographs*, **65**, 283–302.

Pulliam HR (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349–361.

Pulliam RH (1988) Sources, sinks, and population regulation. *The American Naturalist*, **132**, 652–661.

Pybus OG, Suchard MA, Lemey P *et al.* (2012) Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 15066–15071.

Pyron RA, Burbrink FT (2013) Phylogenetic estimates of speciation and extinction rates for testing ecological and evolutionary hypotheses. *Trends in Ecology & Evolution*, **28**, 729–736.

Qian H, Ricklefs RE (2000) Large-scale processes and the Asian bias in species diversity of temperate plants. *Nature*, **407**, 180–182.

Qian H, Swenson NG, Zhang J (2013) Phylogenetic beta diversity of angiosperms in North America. *Global Ecology and Biogeography*, **22**, 1152–1161.

Rabosky DL (2009) Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. *Ecology Letters*, **12**, 735–743.

Rahbek C (2005) The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters*, **8**, 224–239.

Rahbek C, Gotelli NJ, Colwell RK *et al.* (2007) Predicting continental-scale patterns of bird species richness with spatially explicit models. *Proceedings. Biological sciences / The Royal Society*, **274**, 165–174.

Rahbek C, Graves GR (2001) Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4534–4539.

Rangel TFLVB, Diniz-Filho JAF, Colwell RK (2007) Species richness and evolutionary niche dynamics: a spatial pattern-oriented simulation experiment. *The American Naturalist*, **170**, 602–616.

Räsänen K, Hendry AP (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters*, **11**, 624–36.

Reis M Dos, Donoghue PCJ, Yang Z (2014) Neither phylogenomic nor palaeontological data support a Palaeogene origin of placental mammals. *Biology Letters*, **10**, 20131003.

Richards CL, Carstens BC, Knowles L (2007) Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography*, **34**, 1833–1845.

Ricklefs RE (2004) A comprehensive framework for global patterns in biodiversity. *Ecology Letters*, **7**, 1–15.

Ricklefs RE, Jenkins DG (2011) Biogeography and ecology: towards the integration of two disciplines. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **366**, 2438–2448.

Rohde K (1992) Latitudinal gradients in species diversity : the search for the primary cause. *Oikos*, **65**, 514–527.

Rohde K (1996) Rapoport's rule is a local phenomenon and cannot explain latitudinal gradients in species diversity. *Biodiversity Letters*, **3**, 10-13.

Ronquist F, Sanmartín I (2011) Phylogenetic methods in biogeography. *Annual Review of Ecology, Evolution, and Systematics*, **42**, 441–464.

Roquet C, Boucher FC, Thuiller W, Lavergne S (2013a) Replicated radiations of the alpine genus Androsace (Primulaceae) driven by range expansion and convergent key innovations. *Journal of biogeography*, **40**, 1874–1886.

Roquet C, Thuiller W, Lavergne S (2013b) Building megaphylogenies for macroecology: taking up the challenge. *Ecography*, **36**, 13–26.

Rosenzweig ML (1995) *Species diversity in space and time*. Cambridge University Press.

Ruggiero A, Hawkins BA (2008) Why do mountains support so many species of birds ? *Ecography*, **31**, 306–315.

Ruggiero A, Kitzberger T (2004) Environmental correlates of mammal species richness in South America : effects of spatial structure, taxonomy and geographic range. *Ecography*, **27**, 401–416.

Rull V (2009) Microrefugia. *Journal of Biogeography*, **36**, 481–484.

Salisbury CL, Seddon N, Cooney CR, Tobias JA (2012) The latitudinal gradient in dispersal constraints: ecological specialisation drives diversification in tropical birds. *Ecology Letters*, **15**, 847–855.

Schemske DW, Mittelbach GG, Cornell H V., Sobel JM, Roy K (2009) Is there a latitudinal gradient in the importance of biotic interactions? *Annual Review of Ecology, Evolution, and Systematics*, **40**, 245–269.

Schleuter D, Daufresne M, Massol F, Argillier C (2010) A user's guide to functional diversity indices. *Ecological Monographs*, **80**, 469–484.

Schlumberger O, Sagliocco M, Proteau JP (2001) Biogéographie su silure glane (Silurus ganis): causes hydrographiques, climatiques et anthropiques. *Bulletin Français de la Pêche et de la Pisciculture*, **357/360**, 533–547.

Schmitt T (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in zoology*, **4**, 11.

Santucci F, Emerson BC, Hewitt GM (1998) Mitochondrial DNA phylogeography of European hedgehogs. *Molecular Ecology*, **7**, 1163–1172.

Shmida A, Wilson M V (1985) Biological determinants of species diversity. *Journal of Biogeography*, **12**, 1–20.

Simon KS, Benfield EF, Macko SA (2003) Food web structure and the role of epilithic biofilms in cave streams. *Ecology*, **84**, 2395–2406.

Simpson GG (1951) The species concept. *Evolution*, **5**, 285-298.

Sket B (1981) *Niphargobates orophobata* n. gen. n. sp. (Amphipoda *Gammaridae* s.l.) from caves waters in Slovenia (NW Yugoslavia). Bioloski Vestnik, **29**, 105–118.

Smith EP, van Belle G (198AD) Non parametric estimation of species richness. *Biometrics*, **40**, 119–129.

Soberón J, Jiménez R, Golubov J, Koleff P (2007) Assessing completeness of biodiversity databases at different spatial scales. *Ecography*, **30**, 152–160.

Stephenson NL (1998) Actual evapotranspiration and deficit : biologically meaningful correlates of vegetation distribution across spatial scales. *Journal of Biogeography*, **25**, 855–870.

Stevens G (1989) The latitudinal gradient in geographical range: how so many species coexist in the tropics. *The American Naturalist*, **133**, 240–256.

Stewart JR, Lister AM, Barnes I, Dalén L (2010) Refugia revisited: individualistic responses of species in space and time. . *Proceedings of the Royal Society B-Biological Sciences*, **277**, 661–671.

Stoch F, Galassi DMP (2010) Stygobiotic crustacean species richness: a question of numbers, a matter of scale. *Hydrobiologia*, **653**, 217–234.

Stoch F (1995) The ecological and historical determinants of crustacean diversity in groundwaters, or: why are there so many species? *Mémoires de Biospéologie*, **22**, 139-160.

Storch D, Davies RG, Zajícek S *et al.* (2006) Energy, range dynamics and global species richness patterns: reconciling mid-domain effects and environmental determinants of avian diversity. *Ecology Letters*, **9**, 1308–1320.

Struwe L, Smouse PE, Heiberg E, Haag S, Lathrop RG (2011) Spatial evolutionary and ecological vicariance analysis (SEEVA), a novel approach to biogeography and speciation research, with an example from Brazilian Gentianaceae. *Journal of Biogeography*, **38**, 1841–1854.

Sunday JM, Bates AE, Dulvy NK (2011) Global analysis of thermal tolerance and latitude in ectotherms. *Proceedings. Biological sciences / The Royal Society*, **278**, 1823–1830.

Svenning J-C, Fløjgaard C, Baselga A (2011) Climate, history and neutrality as drivers of mammal beta diversity in Europe: insights from multiscale deconstruction. *The Journal of animal ecology*, **80**, 393–402.

Svenning J-C, Normand S, Kageyama M (2008) Glacial refugia of temperate trees in Europe: insights from species distribution modelling. *Journal of Ecology*, **96**, 1117–1127.

Svenning J-C, Normand S, Skov F (2009) Plio-Pleistocene climate change and geographic heterogeneity in plant diversity-environment relationships. *Ecography*, **32**, 13–21.

Svenning J-C, Skov F (2004) Limited filling of the potential range in European tree species. *Ecology Letters*, **7**, 565–573.

Svenning J-C, Skov F (2007) Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? *Ecology letters*, **10**, 453–460.

Swenson NG (2011) Phylogenetic beta diversity metrics, trait evolution and inferring the functional beta diversity of communities. *PloS one*, **6**, e21264.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cossons JF (1998) Comparative phylogeography and postglacial colonization. *Molecular Ecology*, 7, 453–464.

Thuiller W, Münkemüller T, Lavergne S *et al.* (2013) A road map for integrating eco-evolutionary processes into biodiversity models. *Ecology letters*, **16**, 94–105.

Tisseuil C, Cornu J-F, Beauchard O *et al.* (2013) Global diversity patterns and cross-taxa convergence in freshwater systems. *The Journal of animal ecology*, **82**, 365–376.

Toucanne S, Zaragosi S, Bourillet JF *et al.* (2009a) Timing of massive "Fleuve Manche" discharges over the last 350 kyr: insights into the European ice-sheet oscillations and the European drainage network from MIS 10 to 2. *Quaternary Science Reviews*, **28**, 1238–1256.

Toucanne S, Zaragosi S, Bourillet JF *et al.* (2009b) A 1.2 Ma record of glaciation and fluvial discharge from the West European Atlantic margin. *Quaternary Science Reviews*, **28**, 2974–2981.

Trabucco, A. & Zomer, R.J. (2010) Global soil water balance geospatial database. CGIAR Consortium for Spatial Information. Published online. GeoPortal : http://www.cgiar-csi.org .

Trajano E, Bichuette ME (2010) Diversity of Brazilian subterranean invertebrates, with a list of troglomorphic taxa. *Subterranean Biology*, **7**, 1–16.

Trontelj P, Douady CJ, Fišer C *et al.* (2009) A molecular test for cryptic diversity in ground water: how large are the ranges of macro-stygobionts? *Freshwater Biology*, **54**, 727–744.

Trontelj P, Fišer C (2009) Cryptic species diversity should not be trivialised. *Systematics and Biodiversity*, **7**, 1–3.

Tuekam Kayo R, Marmonier P, Zebaze Togouet SH, Nola M, Piscart C (2012) An annotated checklist of freshwater stygobiotic crustaceans of Africa and Madagascar. *Crustaceana,* **85**, 1613-1631.

Tuomisto H (2010a) A diversity of beta diversities: straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. *Ecography*, **33**, 23–45.

Tuomisto H (2010b) A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography*, **33**, 2–22.

Urban MC, Leibold M a, Amarasekare P *et al.* (2008) The evolutionary ecology of metacommunities. *Trends in Ecology & Evolution*, **23**, 311–317.

Valentine, J.M. (1932) Horologion, a new genus of cave beetles (fam. Carabidae). *Annals of the Entomological Society of America*, **25**, 1–8.

Vellend M (2010) Conceptual synthesis in community ecology. *The Quarterly Review of Biology*, **85**, 183–206.

Veter NM, Desantis LRG, Yann LT *et al.* (2013) Is Rapoport's rule a recent phenomenon ? A deep time perspective on potential causal mechanisms. *Biology Letters*, **9**, 20130398.

Voss C, Currell M, Gleeson T (2014) Editors' Message: The 2013 Editors' Choice articles, a new editor, and the 2013 "Coolest Paper" award. *Hydrogeology Journal*, **22**, 293–294.

Waide RB, Willig MR, Steiner CF *et al.* (1999) The relationship between productivity and species richness. *Annual Review of Ecology and Systematics,* **30**, 257–300.

Wakeley J (2009) *Coalescent theory: an introduction*. Robert & Company Publishers, Greenwood Village, Colorado.

Wallace AR (1878) *Tropical nature and other essays*. Macmillan, New York.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular ecology*, **15**, 1419–1439.

Ward J V, Palmer MA (1994) Distribution patterns of interstitial freshwater meiofauna over a range of spatial scales , with emphasis on alluvial river-aquifer systems. *Hydrobiologia*, **287**, 147–156.

Webb C (2000) Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *The American Naturalist*, **156**, 145–155.

Webb CO, Ackerly DD, Kembel SW (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, **24**, 2098–2100.

Webb CO, Ackerly DD, McPeek MA, Donoghue MJ (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, **33**, 475–505.

Weir JT, Schluter D (2007) The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science*, **315**, 1574–1576.

Weiser MD, Enquist BJ, Boyle B *et al.* (2007) Latitudinal patterns of range size and species richness of New World woody plants. *Global Ecology and Biogeography*, **16**, 679–688.

Whittaker RJ, Willis KJ, Field R (2001) Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, **28**, 453–470.

Wiens JJ (2004) Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. *Evolution*, **58**, 193–197.

Wiens JJ (2011) The causes of species richness patterns across space and time, and clades and the role of "ecological limits". *The Quarterly Review of Biology*, **86**, 75–96.

Wiens JJ, Donoghue MJ (2004) Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*, **19**, 639–644.

Willig MR, Kaufman DM, Stevens RD (2003) Latitudinal gradients of biodiversity: Pattern, Process, Scale, and Synthesis. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 273–309.

Wilson EO, Ho B (2005) Eusociality: origin and consequences. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 13367–13371.

Wisz MS, Pottier J, Kissling WD *et al.* (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological reviews of the Cambridge Philosophical Society*, **88**, 15–30.

Wright DH (1983) Species–energy theory, an extension of species–area theory. *Oikos*, **41**, 496–506.

Yang W, Ma K, Kreft H (2013) Geographical sampling bias in a large distributional database and its effects on species richness-environment models (W Daniel Kissling, Ed,). *Journal of Biogeography*, **40**, 1415–1426.

Yang Z, Rannala B (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 9264-9269.

Zagmajster M, Culver DC, Christman MC, Sket B (2010) Evaluating the sampling bias in pattern of subterranean species richness: combining approaches. *Biodiversity and Conservation*, **19**, 3035–3048.

Zagmajster M, Culver DC, Sket B (2008) Species richness patterns of obligate subterranean beetles (Insecta: Coleoptera) in a global biodiversity hotspot - effect of scale and sampling intensity. *Diversity and Distributions*, **14**, 95–105.

Zakšek V, Sket B, Gottstein S, Franjevic D, Trontelj P (2009) The limits of cryptic diversity in groundwater : phylogeography of the cave shrimp *Troglocaris anophthalmus* (Crustacea: Decapoda: Atyidae). *Molecular Ecology*, **18**, 931–946.

Zakšek V, Sket B, Trontelj P (2007) Phylogeny of the cave shrimp *Troglocaris* : Evidence of a young connection between Balkans and Caucasus. *Molecular Phylogenetics and Evolution*, **42**, 223–235.

Zapata FA, Gaston KJ, Chown SL (2005) The Mid-Domain effect revisited. *The American Naturalist*, **166**, E144–E148.

Annexe 1 : Tables des variables environnementales.

Variable name	Description of the variable	Unit	Present/Historical variable	Sources
Slopmw	Mean slope of the river segment	Degree	Present	[a]
Elevmw	Mean Elevation of catchment	Metres	Present	[a]
Elevmin	Catchement with minimal elevation	Metres	Present	[a]
Elevmax	Catchment with maximal elevation	Metres	Present	[a]
Elevr	Elevation range	Metres	Present	[a]
P_tempseasmw	Present Mean temperature seasonality	Degree Celsius	Present	[b]
L_tseasccsmmw	Mean temperature seasonality inferred by CCSM model	Degree Celsius	Historical	[b]
L_tseasmirmw	Mean temperature seasonality inferred by MIROC model	Degree Celsius	Historical	[b]
L_tseasaver	Mean temperature seasonality averaged between CCSM and MIROC models	Degree Celsius	Historical	[b]
Ano_tseasoaver	Anomaly of temperature seasonality averaged between CCSM and MIROC model	Degree Celsius	Stability	[b]
P_solradmw	Mean annual solar radiation	mm (precipitation equivalent / day)	Present	[c]
P_precmw	Mean annual precipitation	Mm	Present	[b]
L_precccsmw	Mean annual precipitation inferred by CCSM model	Mm	Historical	[b]
L_precmirmw	Mean annual precipitation inferred by MIROC model	Mm	Historical	[b]
L_precaver	Mean annual precipitation averaged between CCSM and MIROC models	Mm	Historical	[b]
Ano_precaver	Anomaly of Mean annual precipitation averaged between CCSM and MIROC models	Mm	Stability	[b]
P_aetmw	Actual evapotranspiration (AET)	Mm	Present	[c]
P_petmw	Potential evapotranspiration (PET)	Mm	Present	[c]
P_Temp	Mean annual temperature	Degree Celsius	Present	[b]
L_TempCCSM	Mean annual temperature inferred by CCSM model	Degree Celsius	Historical	[b]
L_TempMIROC	Mean annual temperature inferred by MIROC model	Degree Celsius	Historical	[b]

AnoTempCCSM	Temperature anomaly between Present and LGM infered by CCSM model	Degree Celsius	Stability	[b]
AnoTempMIROC	Temperature anomaly between Present and LGM infered by MIROC model	Degree Celsius	Stability	[b]
L_TempAVER	Mean annual temperature averaged between CCSM and MIROC model	Degree Celsius	Historical	[b]
Ano_TempAVER	Averaged Temperature anomaly between CCSM and MIROC models	Degree Celsius	Stability	[b]
GW_HABITAT	Surface of Groundwater habitat (all classe of the GHME see Cornu et al. 2013) except non aquiferous rock	km2	Present	[d]
CSHPLP	Areal of Consolidated Rock hight permability large pore size	km2	Present	[d]
CRHPSP	Areal of Consolidated Rock hight permability small pore size	km2	Present	[d]
CRLPLP	Areal of Consolidated Rock Low permability large pore size	km2	Present	[d]
CRLPSP	Areal of Consolidated Rock Low permability small pore size	km2	Present	[d]
CRMPLP	Areal of Consolidated Rock Moderate permability large pore size	km2	Present	[d]
CRMPSP	Areal of Consolidated Rock Moderate permability small pore size	km2	Present	[d]
USHPLP	Areal of Unconsolidated Rock hight permability large pore size	km2	Present	[d]
USHPSP	Areal of Unconsolidated Rock hight permability small pore size	km2	Present	[d]
USLPLP	Areal of Unconsolidated Rock Low permability large pore size	km2	Present	[d]
USLPSP	Areal of Unconsolidated Rock Low permability small pore size	km2	Present	[d]
USMPLP	Areal of Unconsolidated Rock Moderate permability large pore size	km2	Present	[d]
USMPSP	Areal of Unconsolidated Rock Moderate permability small pore size	km2	Present	[d]
Naq_P	Areal of non aquiferous rock	km2	Present	[d]
Water_P	Areal of free water	km2	Present	[d]
Glaciers_P	Areal of Glacier	km2	Present	[d]
Lacust_P	Areal of lacustrine water	km2	Present	[d]
Nbpolyha	Number of patch of groundwater habitat excluding non aquiferous rock	No	Present	[d]
Shannon_HA	Diversity of groundwater habitat excluding non aquifeous rock using Shannon index	No	Present	[d]
Tundra_P	Areal cover by Tundra during the LGM	km2	Historical	[e]
Ice_P	Areal cover by Ice during the LGM	km2	Historical	[e]
Loess_P	Areal cover by Loess deposition during the LGM	km2	Historical	[e]

Sanddune_P	Areal cover by Sand and Dune during the LGM	km2	Historical	[e]
Steppe_P	Areal cover by Steppe during the LGM	km2	Historical	[e]
Taiga_P	Areal cover by Taiga during the LGM	km2	Historical	[e]
Woodedstep_P	Areal cover by wooded steppe during the LGM	km2	Historical	[e]
Xerowoodland_P	Areal cover by xerophytic woods during the LGM	km2	Historical	[e]
FOREST	Areal of forested land	km2	Present	[f]
SHRUB	Areal of lands which shrubs	km2	Present	[f]
OPEN	Areal of open land (grasses, bare rock)	km2	Present	[f]
WETLAND	Areal of wetlands	km2	Present	[f]
HUMANUSEIN	Areal of human altered lands (urban, arable, and agricultural lands)	km2	Present	[f]
Arable_p	Areal of arable land	km2	Present	[f]
Agri_p	Areal of Agricultural land	km2	Present	[f]
Urban_p	Areal of urban	km2	Present	[f]
Aridity	Aridity index = PET-AET (ref: Svenning et al. 2009 Ecography, and Eisehardt et al. 2011 PlosOne)	mm.year-1	Present	[c]
Precwetmomw	Present day Precipitation of the Wettest month (BIO 13 Worldclim variables)	Mm	Present	[b]
Precdrymomw	Present day Precipitation of the Dryest month (BIO 14 Worldclim variables)	Mm	Present	[b]
Precseasomw	Present day precipitation seasonality (BIO 15 worldclim Variables) =coefficient of variation	Mm	Present	[b]
Precwetqamw	Present day Precipitation of the Wettest Quater (BIO 16 Worldclim variables)	Mm	Present	[b]
Precdryqamw	Present day Precipitation of the Dryest Quater (BIO 17 Worldclim variables)	Mm	Present	[b]
Precwarqamw	Present day Precipitation of the Warmest Quater (BIO 18 Worldclim variables)	Mm	Present	[b]
Preccolqamw	Present day Precipitation of the Coldest Quater (BIO 19 Worldclim variables)	Mm	Present	[b]
LandMass501	Summed of available groundwater habitat within a radius of 501 km	km2	Present	[d]
LandMass1001	Summed of available groundwater habitat within a radius of 1001 km	km2	Present	[d]

LandMass1501	Summed of available groundwater habitat within a radius of 1501 km	km2	Present	[d]
ClimaticRarity_301log	Climatic rarity [g]	No	Present	[b]
ClimaticRarity_501log	Climatic rarity [g]	No	Present	[b]
ClimaticRarity_1001log	Climatic rarity [g]	No	Present	[b]
cv_aet	Coefficient of variation of AET computed using 12 month values	Degree Celsius	Present	[b]
P_BIO2	Bioclim variable number 2 either Mean Diurnal Range BIO2	1/10°C	Present	[b]
P_BIO3	Bioclim variable number 3 either Isothermality BIO3	None	Present	[b]
P_BIO4	Bioclim variable number 4 either Temperature Seasonality BIO4	1/1000°C	Present	[b]
P_BIO5	Bioclim variable number 5 either Max Temperature of the Warmest Month x 10 BIO5	1/10°C	Present	[b]
P_BIO6	Bioclim variable number 6 either Max Temperature of the Coldest Month x 10 BIO6	1/10°C	Present	[b]
P_BIO7	Bioclim variable number 7 either Temperature annual Range x 10 BIO7	1/10°C	Present	[b]
P_BIO8	Bioclim variable number 8 either Mean Temperature of Wettest Quarter x 10 BIO8	1/10°C	Present	[b]
P_BIO9	Bioclim variable number 9 either Mean Temperature of Driest Quarter x 10 BIO9	1/10°C	Present	[b]
P_BIO10	Bioclim variable number 10 either Mean Temperature of Warmest Quarter x 10 BIO10	1/10°C	Present	[b]
P_BIO11	Bioclim variable number 11 either Mean Temperature of Coldest Quarter x 10 BIO11	1/10°C	Present	[b]
perm_cont_bv	Continuous Permafrost<=8°C	Nb Cell (Count or Sum) or M² (area)	Present	[b]
perm_dcont_bv	Discontinuous Permafrost Between -8 and -4°C	Nb Cell (Count or Sum) or M² (area)	Present	[b]
perm_isol_bv	Isolated Permafrost, Between -4 and -2°C	Nb Cell (Count or Sum) or M² (area)	Present	[b]
perm_spo_bv	Sporadic Permafrost, Between -2 and 0°C	Nb Cell (Count or Sum) or M² (area)	Present	[b]
L_CCSM_BIO2	Bioclim variable number 2 either Mean Diurnal Range BIO2, CCSM model	1/10°C	Historical	[b]
L_CCSM_BIO3	Bioclim variable number 3 either Isothermality BIO3, CCSM model	None	Historical	[b]

L_CCSM_BIO4	Bioclim variable number 4 either Temperature Seasonality BIO4, CCSM model	1/1000°c	Historical	[b]
L_CCSM_BIO5	Bioclim variable number 5 either Max Temperature of the Warmest Month x 10 BIO5, CCSM model	1/10°C	Historical	[b]
L_CCSM_BIO6	Bioclim variable number 6 either Max Temperature of the Coldest Month x 10 BIO6, CCSM model	1/10°C	Historical	[b]
L_CCSM_BIO7	Bioclim variable number 7 either Temperature annual Range x 10 BIO7, CCSM model	1/10°C	Historical	[b]
L_CCSM_BIO8	Bioclim variable number 8 either Mean Temperature of Wettest Quarter x 10 BIO8, CCSM model	1/10°C	Historical	[b]
L_CCSM_BIO9	Bioclim variable number 9 either Mean Temperature of Driest Quarter x 10 BIO9, CCSM model	1/10°C	Historical	[b]
L_CCSM_BIO10	Bioclim variable number 10 either Mean Temperature of Warmest Quarter x 10 BIO10, CCSM model	1/10°C	Historical	[b]
L_CCSM_BIO11	Bioclim variable number 11 either Mean Temperature of Coldest Quarter x 10 BIO11, CCSM model	1/10°C	Historical	[b]
L_CCSM_BIO12	Bioclim variable number 13 either Precipitation of Wettest Month BIO13, CCSM model	Mm	Historical	[b]
L_CCSM_BIO13	Bioclim variable number 14 either Precipitation of Driest Month BIO14, CCSM model	Mm	Historical	[b]
L_CCSM_BIO14	Bioclim variable number 15 either Precipitation Seasonality BIO15, CCSM model	Mm	Historical	[b]
L_CCSM_BIO15	Bioclim variable number 16 either Precipitation of Wettest Quarter BIO16, CCSM model	Mm	Historical	[b]
L_CCSM_BIO16	Bioclim variable number 17 either Precipitation of Driest Quarter BIO17, CCSM model	Mm	Historical	[b]
L_CCSM_BIO17	Bioclim variable number 18 either Precipitation of Warmest Quarter BIO18, CCSM model	Mm	Historical	[b]
L_CCSM_BIO18	Bioclim variable number 19 either Precipitation of Coldest Quarter BIO19, CCSM model	Mm	Historical	[b]
L_CCSM_perm_cont	Continuous Permafrost<=8°C , CCSM model	Nb Cell (Count or Sum) or M² (area)	Historical	[b]
L_CCSM_perm_dcont	Discontinuous Permafrost Between -8 and -4°C, CCSM model	Nb Cell (Count or Sum) or M² (area)	Historical	[b]

L_CCSM_perm_isol	Isolated Permafrost, Between -4 and -2°C, CCSM model	Nb Cell (Count or Sum) or M² (area)	Historical	b
L_CCSM_perm_spo	Sporadic Permafrost, Between -2 and 0°C, CCSM model	Nb Cell (Count or Sum) or M² (area)	Historical	b
L_MIROC3_BIO2	Bioclim variable number 2 either Mean Diurnal Range BIO2, MIROC3 model	1/10°C	Historical	b
L_MIROC3_BIO3	Bioclim variable number 3 either Isothermality BIO3, MIROC3 model	None	Historical	b
L_MIROC3_BIO4	Bioclim variable number 4 either Temperature Seasonality BIO4, MIROC3 model	1/1000°C	Historical	b
L_MIROC3_BIO5	Bioclim variable number 5 either Max Temperature of the Warmest Month x 10 BIO5, MIROC3 model	1/10°C	Historical	b
L_MIROC3_BIO6	Bioclim variable number 6 either Max Temperature of the Coldest Month x 10 BIO6, MIROC3 model	1/10°C	Historical	b
L_MIROC3_BIO7	Bioclim variable number 7 either Temperature annual Range x 10 BIO7, MIROC3 model	1/10°C	Historical	b
L_MIROC3_BIO8	Bioclim variable number 8 either Mean Temperature of Wettest Quarter x 10 BIO8, MIROC3 model	1/10°C	Historical	b
L_MIROC3_BIO9	Bioclim variable number 9 either Mean Temperature of Driest Quarter x 10 BIO9, MIROC3 model	1/10°C	Historical	b
L_MIROC3_BIO10	Bioclim variable number 10 either Mean Temperature of Warmest Quarter x 10 BIO10, MIROC3 model	1/10°C	Historical	b
L_MIROC3_BIO11	Bioclim variable number 11 either Mean Temperature of Coldest Quarter x 10 BIO11, MIROC3 model	1/10°C	Historical	b
L_MIROC3_BIO12	Bioclim variable number 13 either Precipitation of Wettest Month BIO13, MIROC3 model	Mm	Historical	b
L_MIROC3_BIO13	Bioclim variable number 14 either Precipitation of Driest Month BIO14, MIROC3 model	Mm	Historical	b
L_MIROC3_BIO14	Bioclim variable number 15 either Precipitation Seasonality BIO15, MIROC3 model	Mm	Historical	b
L_MIROC3_BIO15	Bioclim variable number 16 either Precipitation of Wettest Quarter BIO16, MIROC3 model	Mm	Historical	b
L_MIROC3_BIO16	Bioclim variable number 17 either Precipitation of Driest Quarter BIO17, MIROC3 model	Mm	Historical	b

L_MIROC3_BIO17	Bioclim variable number 18 either Precipitation of Warmest Quarter BIO18, MIROC3 model	Mm	Historical	[b]
L_MIROC3_BIO18	Bioclim variable number 19 either Precipitation of Coldest Quarter BIO19, MIROC3 model	Mm	Historical	[b]
L_MIROC3_perm_cont	Continuous Permafrost<=8°C , MIROC3 model	Nb Cell (Count or Sum) or M² (area)	Historical	[b]
L_MIROC3_perm_dcont	Discontinuous Permafrost Between -8 and -4°C, MIROC3 model	Nb Cell (Count or Sum) or M² (area)	Historical	[b]
L_MIROC3_perm_isol	Isolated Permafrost, Between -4 and -2°C, MIROC3 model	Nb Cell (Count or Sum) or M² (area)	Historical	[b]
L_MIROC3_perm_spo	Sporadic Permafrost, Between -2 and 0°C, MIROC3 model	Nb Cell (Count or Sum) or M² (area)	Historical	[b]

[a] Voght et al. 2007, Catchment Characterisation and Modeling (CCM2), http://ccm.jrc.ec.europa.eu/php/index.php?action=view&id=23

[b] Hijmans et al. 2005, WorldClim, http://www.worldclim.org/

[c] Zomer & Trabucco, 2010, Consortium for Spatial Information CGIAR-CSI, http://www.cgiar-csi.org/data/item/60-global-high-resolution-soil-water-balance

[d] Cornu et al. 2013, BioFresh data portal, http://data.freshwaterbiodiversity.eu/data/shapefiles/

[e] Ehlers et al. 2004 ; Hughes Woodward 2008 ; Commision for the Geological Map of the World. 1999. Maps of the world environment during the last two climatic extremes. The last glacial maximum.

[f] GlobeCover, (2009) European Space Agengy ESA, http://due.esrin.esa.int/globcover/

[g] Climatic rarity computed using averaged euclidian distance between focal cell and neighbors cells (binary coding scheme for the neighbors) within a radius of 1001km using the coordinates of the two first axis of a normalized PCA computed using seven Wordclim variables (BIO1, 12,13,14,16,17,18,19), log transform to comply with the normality assumption (Morueta -Holme et al. 2013).

Annexe 2: Matériel supplémentaire Article 1

The distribution of groundwater habitats in Europe. Jean-François Cornu, David Eme and Florian Malard. Electronic supplementary material. Hydrogeology Journal.

Figure ESM1: Mosaic of the 25 map sheets of the international hydrogeological map of Europe (IHME). a: original scanned map sheets (img format). b: vector IHME. Colors in (a) and (b) correspond to the second classification level of the IHME referred to as aquifer productivity.

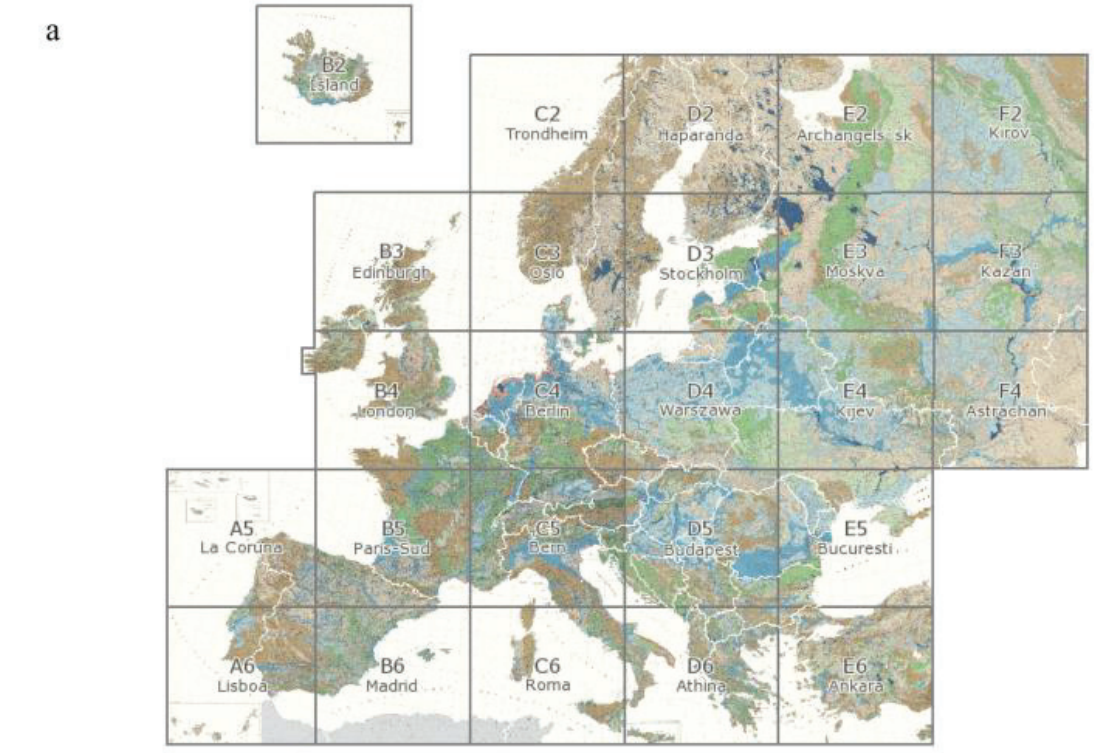


Figure ESM2: Distribution of catchments (black lines; n=778) used to test for latitudinal variation in habitat diversity and biogeographical regions (red lines; n=22) used to test for the effect of habitat diversity on regional groundwater species richness. Regions names are given in Table ESM1. Colors show the residuals of the relationships between habitat diversity and catchment area (see Figure 5b of the main article).

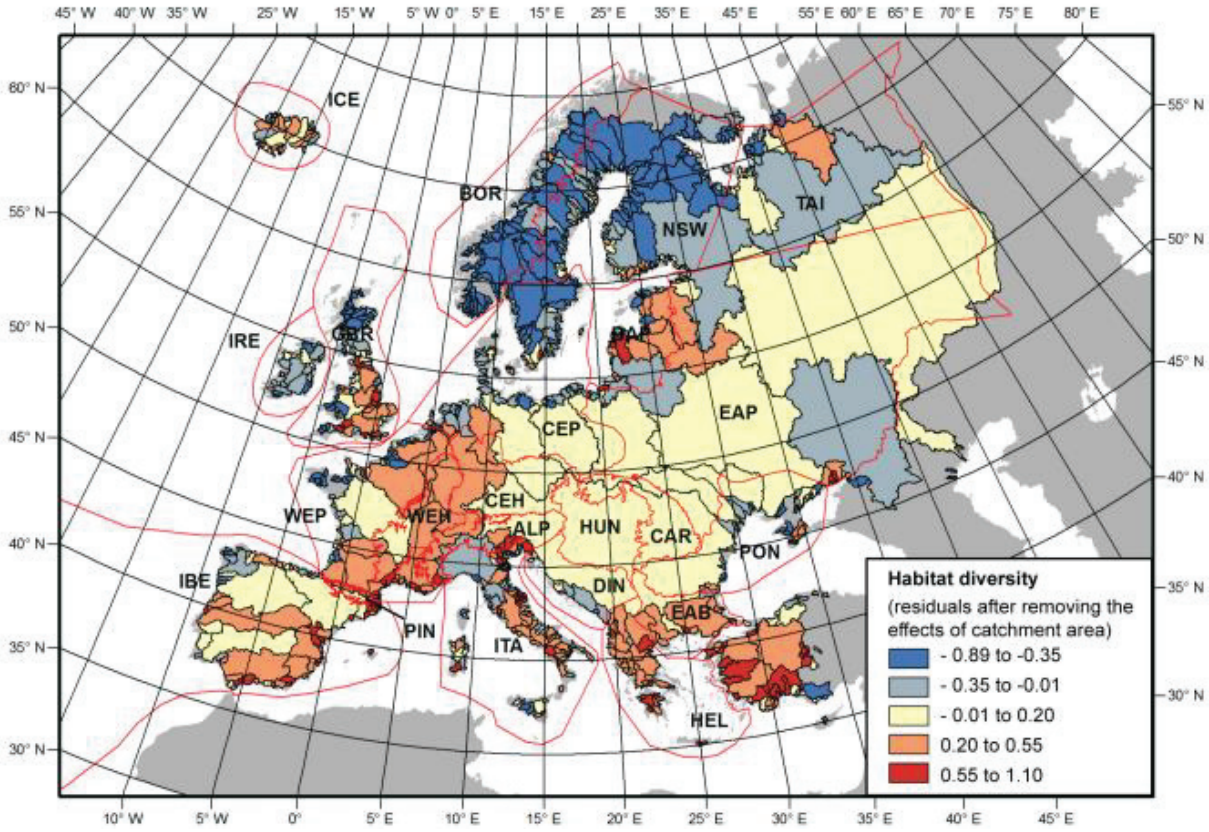


Table ESM1: Data used in the generalized linear model of regional groundwater species richness. See Figure ESM2 for the spatial distribution of biogeographical regions. Data for longitude, latitude, maximum elevation, area and groundwater species richness are from Hof et al. (2008).

Biogeographical regions	Region code	Longitude centre (°N)	Latitude centre (°N)	Maximum elevation (m)	Area (km^2)	Habitat diversity	Groundwater species richness
Borealic uplands	BOR	12.75	64.68	2469	338 784	0.78	4
Fenno-scandian shield	NSW	26.37	64.40	1200	698 004	1.02	12
Ireland and Northern Ireland	IRE	- 8.27	53.46	1041	80 638	1.15	9
Pyrenees	PIN	0.79	42.63	3404	18 261	1.46	82
Dinaric western Balkan	DIN	17.36	44.05	2656	144 468	1.58	170
England	GBR	- 3.69	56.39	1344	226 892	1.60	33
Iceland	ICE	- 18.97	65.07	2110	92 025	1.65	2
Alps	ALP	10.73	46.46	4808	148 998	1.72	96
The Carpathiens	CAR	23.16	47.37	2655	156 328	1.78	130
Italy and Corsica	ITA	12.43	40.83	3380	269 745	1.80	135
Taiga	TAI	47.90	62.43	1000	921 654	1.81	9
Central plains	CEP	13.40	55.05	500	525 002	1.83	77
Hungarian lowlands	HUN	19.43	47.05	500	214 146	1.88	67
Western highlands	WEH	5.16	47.17	1886	169 673	1.89	101
Hellenic western Balkan	HEL	23.17	38.09	2917	161 565	1.95	100
Eastern Balkan	EAB	24.81	41.77	2925	145 668	1.99	74
Central highlands	CEH	11.53	49.60	1602	252 443	2.07	133
Ibero-Macaronesian region	IBE	- 14.66	37.00	3717	584 000	2.08	99
Western plains	WEP	0.91	47.28	1000	404 026	2.15	106
Eastern plains	EAP	38.17	53.83	1000	2 243 785	2.16	27
Pontic province	PON	30.69	45.55	1545	298 954	2.17	50
Baltic province	BAP	23.48	57.13	318	178 367	2.21	13

Annexe 3 : Matériel supplémentaire Article 2

Appendix S1: Relationships between median range size or species richness and latitude.

Table S1.1: Results of ordinary least square and generalized additive models used to test for relationships between species richness (SR), maximal linear range extent (MLE), area of occupancy (AOO) and latitude. Slopes and smoothing parameters (Edf) were all significant ($p < 0.0001$).

Variable	n [a]	Ordinary least square model					Generalized additive model			
		Slope	Standard error	t	R^2_{adj}	AIC [b]	Edf [c]	F	Explained deviance (%)	AIC
SR	40	- 0.34	0.07	-5.6	0.44	235.8	8.16	28.6	89.2	185.4
MLE	36 [d]	113.50	6.56	17.3	0.90	531.8	7.72	215.6	98.6	475.2
AOO	36	5.26	0.29	18.3	0.91	306.8	6.49	83.5	95.7	290.2

[a] Number of latitudinal bands

[b] Akaike information criterion

[c] effective degree of freedom

[d] $n = 36$ bands for median range size to allow comparison with results of generalized least square models in Table S1.2 (see below)

Table S1.2: Comparison of cell average species richness, maximum linear extent and area of occupancy among latitudinal bands. All bands are compared to a single reference band (42.75°N; intercept). Differences in species richness of cells among latitudinal bands were tested by means of negative binomial generalized linear models to account for overdispersion. Differences in range size were tested by means of generalized least squares, while accounting for unequal variances among latitudinal bands. In all models, *p* values were adjusted with Holm corrections for multiple comparisons. Significant *p* values are indicated in bold.

	Species richness				Maximum linear extent				Area of occupancy			
Latitudinal band (°N)	Estimate	Standard Error	z value	*p*	Estimate	Standard Error	t value	*p*	Estimate	Standard Error	t value	*p*
42.75	3.27	0.16	20.8	**1.67E-94**	235.8	95.9	2.46	1.57E-01	6.21	0.96	6.45	**6.87E-09**
35.55	-1.54	0.52	-2.96	**1.82E-02**	304.6	287.6	1.06	9.39E-01	1.12	3.61	0.31	9.47E-01
36.45	-1.32	0.51	-2.61	**3.61E-02**	-80.5	287.6	-0.28	9.39E-01	-2.87	2.30	-1.25	9.47E-01
37.35	-1.09	0.28	-3.95	**6.94E-04**	326.8	161.7	2.02	3.96E-01	2.56	3.27	0.78	9.47E-01
38.25	-1.54	0.27	-5.81	**1.67E-07**	283.4	151.6	1.87	4.98E-01	9.51	6.72	1.41	9.47E-01
39.15	-1.14	0.28	-4.1	**4.88E-04**	187.9	161.7	1.16	9.39E-01	4.91	3.61	1.36	9.47E-01
40.05	-1.07	0.25	-4.3	**2.42E-04**	174.2	146.4	1.19	9.39E-01	3.49	2.90	1.20	9.47E-01
40.95	-1.07	0.24	-4.5	**1.24E-04**	253.7	140.3	1.81	4.99E-01	3.36	2.10	1.60	8.85E-01
41.85	-0.69	0.23	-2.95	**1.82E-02**	228.7	140.3	1.63	6.24E-01	4.01	2.62	1.53	8.92E-01
43.65	-0.36	0.22	-1.64	3.05E-01	179.4	132.9	1.35	8.90E-01	6.04	2.55	2.37	1.63E-01
44.55	-0.74	0.23	-3.25	**8.08E-03**	321.9	135.6	2.37	1.80E-01	14.85	3.78	3.93	**1.29E-03**
45.45	-0.18	0.22	-0.81	4.15E-01	415.8	135.6	3.07	**2.76E-02**	19.02	5.09	3.73	**2.55E-03**
46.35	-0.3	0.23	-1.33	3.64E-01	432.3	137.0	3.15	**2.24E-02**	13.16	3.20	4.11	**6.99E-04**
47.25	-0.92	0.23	-4.09	**4.88E-04**	723.4	134.2	5.39	**1.63E-06**	37.79	9.24	4.09	**7.22E-04**
48.15	-0.85	0.22	-3.79	**1.21E-03**	760.5	134.2	5.66	**4.01E-07**	40.61	5.98	6.80	**9.24E-10**
49.05	-1.02	0.24	-4.29	**2.42E-04**	1060.3	140.3	7.55	**4.52E-12**	63.72	9.81	6.50	**5,39E-09**
49.95	-1.35	0.25	-5.43	**1.20E-06**	1236.4	144.2	8.57	**3.61E-15**	58.63	9.33	6.29	**1.74E-08**
50.85	-1.33	0.24	-5.53	**7.09E-07**	1196.7	140.3	8.52	**4.82E-15**	60.60	8.54	7.10	**1.43E-10**
51.75	-1.49	0.26	-5.63	**4.29E-07**	1347.9	151.6	8.89	**3.43E-16**	59.73	13.86	4.31	**3.66E-04**
52.65	-2.17	0.31	-7.12	**3.99E-11**	1162.0	161.7	7.18	**4.77E-11**	60.68	14.64	4.14	**6.54E-04**
53.55	-2.71	0.41	-6.66	**9.13E-10**	1707.9	191.7	8.91	**3.21E-16**	90.42	32.80	2.76	6.08E-02
54.45	-1.95	0.48	-4.06	**4.88E-04**	1897.3	253.7	7.48	**7.11E-12**	33.54	2.83	11.87	**7.90E-27**
55.35	-2.58	0.45	-5.75	**2.25E-07**	2681.2	214.4	12.50	**1.96E-29**	95.62	18.65	5.13	**8.48E-06**
56.25	-2.44	0.41	-5.99	**6.34E-08**	2548.4	201.7	12.63	**6.67E-30**	67.86	18.57	3.65	**3.18E-03**
57.15	-2.66	0.46	-5.84	**1.46E-07**	2676.9	214.4	12.48	**2.28E-29**	104.87	19.11	5.49	**1.45E-06**
58.05	-2.58	0.54	-4.8	**3.04E-05**	2452.6	253.7	9.67	**8.57E-19**	105.92	19.93	5.31	**3.43E-06**
58.95	-2.71	0.41	-6.66	**9.13E-10**	2876.8	191.7	15.00	**9.47E-40**	132.23	8.83	14.97	**1.15E-39**
59.85	-2.76	0.47	-5.92	**9.28E-08**	2933.7	214.4	13.68	**3.34E-34**	132.79	6.57	20.21	**5.51E-63**
60.75	-2.87	0.42	-6.8	**3.62E-10**	2833.9	191.7	14.78	**8.22E-39**	125.17	17.88	7.00	**2.66E-10**
61.65	-2.9	0.4	-7.18	**2.60E-11**	2750.1	183.6	14.98	**1.16E-39**	129.35	15.46	8.37	**2.45E-14**
62.55	-2.93	0.53	-5.56	**6.25E-07**	2944.9	230.9	12.75	**2.18E-30**	135.39	7.36	18.40	**8.51E-55**
63.45	-3.03	0.38	-7.91	**9.91E-14**	2767.3	183.6	15.07	**4.83E-40**	117.96	17.48	6.75	**1.21E-09**
64.35	-2.73	0.43	-6.31	**8.62E-09**	2911.4	230.9	12.61	**7.87E-30**	147.09	4.87	30.19	**3.20E-107**
65.25	-2.83	0.4	-7.12	**3.99E-11**	2683.0	230.9	11.62	**5.56E-26**	124.89	17.18	7.27	**4.92E-11**
66.15	-2.71	0.41	-6.66	**9.13E-10**	2860.7	201.7	14.18	**2.88E-36**	137.08	12.56	10.92	**3.39E-23**
67.05	-3.27	0.74	-4.44	**1.54E-04**	2762.8	230.9	11.96	**2.58E-27**	128.39	30.41	4.22	**5.03E-04**
67.95	-3.27	0.58	-5.64	**4.14E-07**								
68.85	-3.27	0.64	-5.09	**7.31E-06**								
69.75	-3.27	0.74	-4.44	**1.54E-04**								
70.65	-3.27	0.74	-4.44	**1.54E-04**								

Appendix S2: Latitudinal patterns of area of occupancy.

Table S2.1: Results of ordinary least square, generalized least square and phylogenetic generalized least square models between species' area of occupancy (Log_{10} (AOO)) and latitudinal midpoint of species' ranges (i.e. mean latitude; n = 1568 species, 2 species endemic to Iceland were excluded). Significance levels * $p < 0.05$, ** $p < 0.01$.

Model	Parameter	Parameter estimate	Standard error	t	R^2	AIC[e]
OLS [a]	Intercept	2. 3	1.2	1.9	0.10	1 950
	Linear	-0.1*	0.1	-2.3		
	Quadratic	0.1**	0.0	3.1		
GLS [b]	Intercept	4.0**	1.4	2.9	0.19	1 838
	Linear	-0.2**	0.1	-3.2		
	Quadratic	0.1**	0.0	3.8		
PGLS [c]	Intercept	2.3	1.2	1.9	0.09	1 891
	Linear	-0.1*	0.1	-2.3		
	Quadratic	0.1**	0.0	3.1		
PGLS [d]	Intercept	3.5*	1.4	2.5	0.19	1 769
	Linear	-0.2**	0.1	-2.8		
	Quadratic	0.1**	0.0	3.3		

[a] Ordinary least squares

[b] Generalized least squares with an exponential variance structure for the residuals

[c] Phylogenetic generalized least squares with the best evolution model (Ornstein-Uhlenbeck)

[d] Phylogenetic generalized least squares with an exponential variance structure for the residuals and the best evolution model (Ornstein-Uhlenbeck)

[e] Akaike information criterion

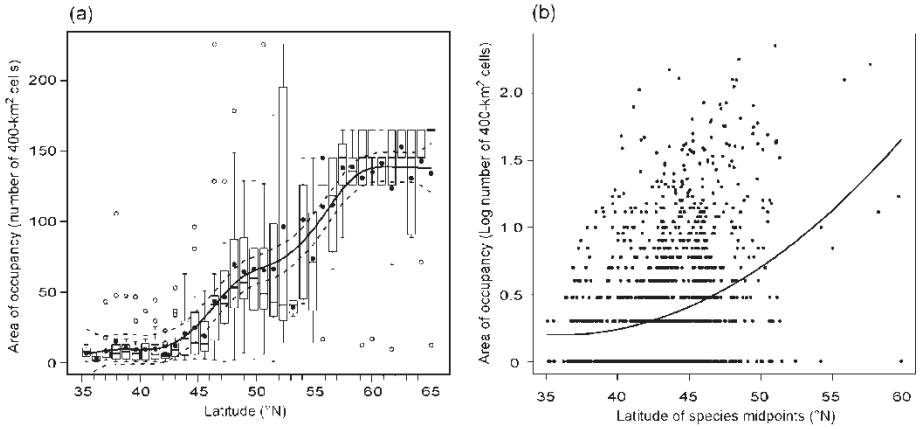
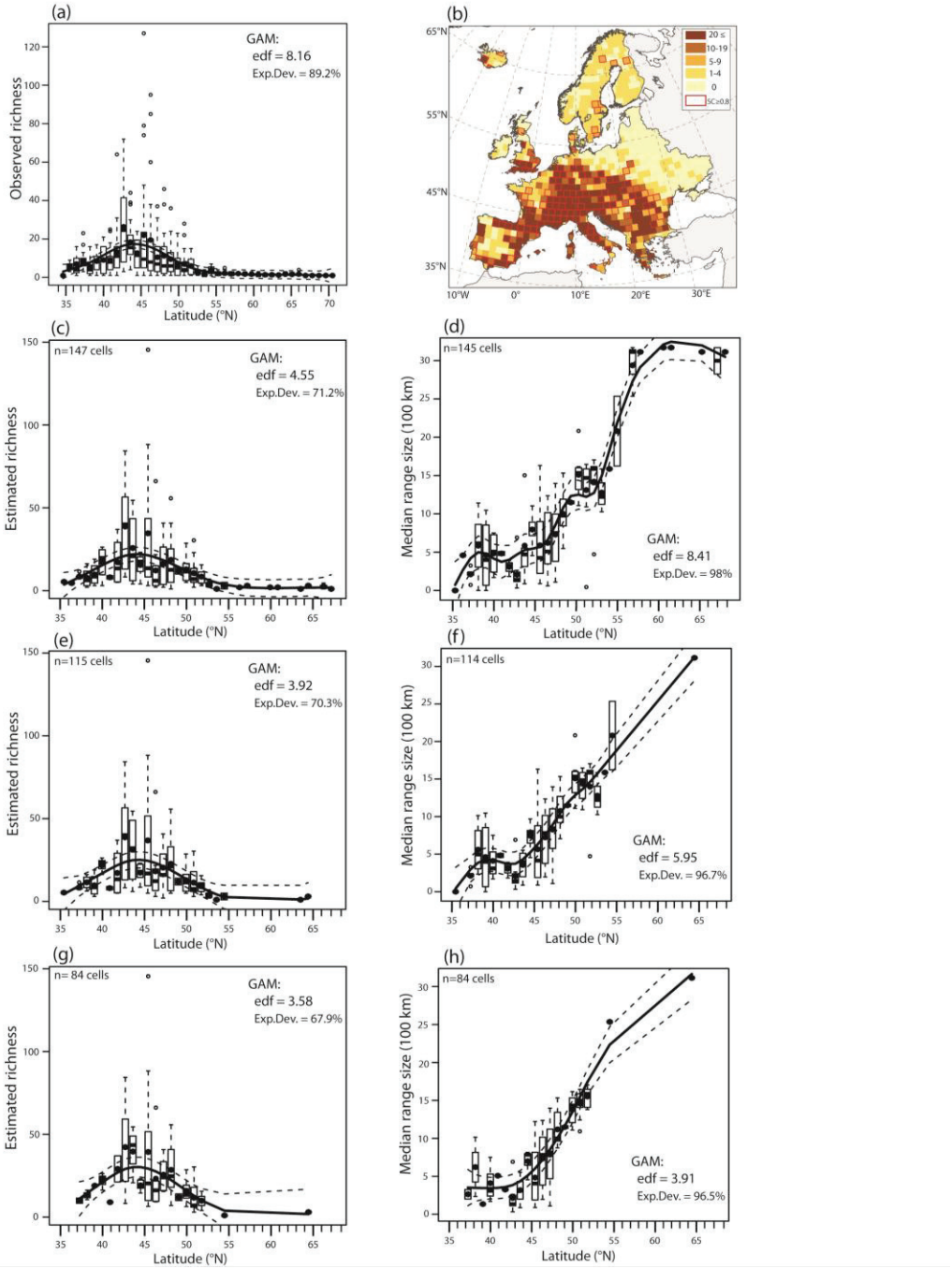


Figure S2.2: Latitudinal patterns of area of occupancy (AOO). (a) Relationship between cell average of area of occupancy per latitudinal band and latitude. Boxplot conventions as in Fig. 2a. The black line shows the fit of a generalized additive model to the averages of latitudinal bands and dashed lines show 95% confidence intervals (see Table S1.1 for statistics). (b) Relationship between area of occupancy of species (n = 1568) and latitude. The black line shows the fit of a phylogenetic generalized least square models using the Ornstein-Uhlenbeck model of evolution, a quadratic term for latitudinal midpoint and an exponential variance structure for the residuals (see Table S2.1 for statistics).

Appendix S3. Patterns of species richness and median range size (maximum linear extent) per 0.9° latitudinal band, when sampling bias is accounted for. Top: observed species richness (a) and map showing the position of cells with at least 5, 10 or 20 records and sampling completeness (SC) > 0.8 (b). Graphs below:

Chao 1 estimated species richness (left side) and median range size (right side) of cells with SC > 0.8 and containing at least 5 (c and d), 10 (e and f) or 20 (g and h) records. Number of cells included in the analyses (n) are given in the respective plots (n = 145 in (d) and n = 114 in (f) because we excluded 2 species endemic to Iceland). Black horizontal bars, black dots and boxes show the median, average and interquartile range, respectively, for latitudinal bands. The maximum length of each whisker is up 1.5 times the interquartile range and open circles represent outliers. Thick black lines represent the fit of generalized additive model (GAM) to the averages of latitudinal bands. Effective degree of freedom (edf; all significant at $p < 0.01$) and the proportion of explained deviance (Exp.Dev.) are provided in each plot.



Appendix S4: Statistical procedure and results of ordinary least squares (OLS) and simultaneous autoregressive (SAR) models used to determine drivers of spatial variation in median range size.

Statistical procedure

To test the three main hypotheses on the drivers of spatial patterns in range size, we performed multiple OLS models using the 7 predictors. Then, variance partitioning (Legendre & Legendre 1998) was used to estimate the unique and shared contributions of the three hypotheses. We ran all possible subsets of the full OLS model and retained only those models whose difference in AICc (AIC corrected for small sample size) with the best model was < 5. The relative importance of each predictor was estimated by computing the sum of AICc weights of models in which the predictor was retained. Variance partitioning for OLS models was computed using the adjusted R^2.

We reiterated the same analytical procedure as described above using simultaneous autoregressive (SAR) models because spatial autocorrelation in the residuals of OLS models could potentially affect the estimation of parameters and their statistical significance (Dormann *et al*., 2007). To select the most appropriate SAR models (i.e. the one with no spatial autocorrelation in the residuals and minimum AICc), we tested a range of neighborhood distances between cells (i.e. 220-2500 km) using a variance stabilizing coding scheme for the spatial weight matrix. The variance explained by the best SAR model was computed using pseudo-R^2 (including the spatial component) whereas the unique and shared contributions of the three hypotheses were calculated using partial-pseudo-R^2 values (excluding the spatial component; Morueta-Holme *et al*., 2013).

Model selection and multi-model inference were performed using MuMIn R package (Barton, 2013). Neighborhood distance matrices and SAR models were computed with the spdep R package (Bivand *et al*., 2012). Variation partitioning was performed with the vegan package (Oksanen *et al*., 2012) for OLS and with a simple set of equations following Legendre & Legendre (1998) for SAR models.

References:

Barton, K. (2013) MuMIn: Multi-model inference. R package version 1.9.0. Available at: http://CRAN.R-project.org/package=MuMIn

Bivand, R., Altman, M., Anselin, L., Assunção, R., Berke, O., Bernat, A., Blanchet, G., Blankmeyer, E., Carvalho, M., Christensen, B., Chun, Y., Dormann, C., Dray, S., Halbersma, R., Krainski, E., Legendre, P., Lewin-Koh, N., Li, H., Ma, J., Millo, G., Mueller, W., Ono, H., Peres-Neto, P., Piras, G., Reder, M., Tiefelsdorf, M. & Yu, D. (2012). spdep: Spatial dependence: weighting schemes, statistics and models. R package version 0.5-53. Available at: http://CRAN.R-project.org/package=spdep

Results of OLS and SAR models

Table S4.1: Full data set (n = 494 cells; OLS: R^2_{adj} = 0.72; SAR: pseudo-R^2 = 0.81). Parameter estimates correspond to averages computed from the subset of models whose delta AICc ≤ 5 compared to the best model (lowest AICc). AICc weights (AICc w) indicate the relative importance of each predictor.

Variable	OLS				SAR			
	Estimate (standard error)	*z value*	*p*	AICc w	Estimate (standard error)	*z value*	*p*	AICc w
intercept	1287.2 (26.0)	49.4	<0.0001	-	1499.5 (87.3)	17.2	<0.0001	-
AET	-198.1 (33.6)	5.9	<0.0001	1	-80.7 (45.7)	1.8	0.07	0.62
Aquifer area	-168.9 (30.1)	5.6	<0.0001	1	-32.9 (34.6)	1	0.34	0.33
Log (Elevation range)	-66.9 (32.9)	2.0	<0.0001	0.79	-43.5 (36.6)	1.2	0.23	0.42
Log (Climatic rarity)	2.2 (39.3)	0.1	0.95	0.25	23.4 (39.4)	0.6	0.55	0.28
Log (Temperature anomaly)	802.6 (40.1)	20.1	<0.0001	1	744.1 (75.1)	9.9	<0.0001	1
Precipitation anomaly	-15.3 (35.9)	0.4	0.67	0.25	105.4 (57.7)	1.8	0.07	0.67
Log (Precipitation seasonality)	64.1 (29.7)	2.2	0.03	0.82	29.9 (47.4)	0.6	0.53	0.28

Table S4.2: Data subset with n = 147 cells (OLS: R^2_{adj} = 0.55; SAR: pseudo-R^2 = 0.73). Parameter estimates and AICc weights as in Table S4.1.

Variable	OLS				SAR			
	Estimate (standard error)	*z value*	*p*	AICc w	Estimate (standard error)	*z value*	*p*	AICc w
intercept	1126.3 (46.8)	23.8	<0.0001	-	1211.9 (89.5)	13.6	<0.0001	-
AET	-83.8 (55.9)	1.5	0.14	0.52	-104.3 (59.3)	1.8	0.08	0.62
Aquifer area	-157.8 (52.7)	3.0	<0.01	1	-64.8 (57.4)	1.1	0.26	0.37
Log (Elevation range)	-97.8 (59.6)	1.6	0.010	0.58	-100.2 (56.3)	1.8	0.08	0.64
Log (Climatic rarity)	-19.8 (70.7)	0.3	0.78	0.24	8.7 (68.2)	0.1	0.90	0.21
Log (Temperature anomaly)	678.4 (74.6)	9.1	<0.0001	1	714.3 (87.3)	8.2	<0.0001	1
Precipitation anomaly	-60.5 (71.3)	0.8	0.40	0.32	-8.9 (88.9)	0.1	0.92	0.20
Log (Precipitation seasonality)	7.8 (47.7)	0.2	0.87	0.24	66.5 (59.7)	1.1	0.27	0.36

Table S4.3: Data subset with n = 115 cells (OLS: R^2_{adj} = 0.49; SAR: pseudo-R^2 = 0.70). Parameter estimates and AICc weights as in Table S4.1.

Variable	OLS				SAR			
	Estimate (standard error)	*z value*	*p*	AICc w	Estimate (standard error)	*z value*	*p*	AICc w
intercept	998.4 (44.7)	22.1	<0.0001	-	875.0 (69.5)	12.6	<0.0001	-
AET	-40.2 (49.3)	0.8	0.42	0.30	-49.0 (46.7)	1.1	0.29	0.33
Aquifer area	-69.1 (45.2)	1.5	0.13	0.52	-74.8 (45.5)	1.6	0.10	0.54
Log (Elevation range)	-100.9 (50.3)	2.0	0.05	0.77	-95.9 (45.2)	2.1	0.04	0.81
Log (Climatic rarity)	43.5 (65.5)	0.7	0.51	0.23	73.1 (61.7)	0.9	0.39	0.35
Log (Temperature anomaly)	488.2 (75.4)	6.2	<0.0001	1	432.2 (81.6)	5.3	<0.0001	1
Precipitation anomaly	-69.00 (62.7)	1.1	0.28	0.35	-60.7 (70.4)	0.9	0.39	0.29
Log (Precipitation seasonality)	-21.7 (39.9)	0.5	0.59	0.23	3.5 (49.7)	0.1	0.94	0.18

Table S4.4: Data subset with n = 83 cells. One isolated cell in Iceland was removed because it had no neighbors. (OLS: R^2_{adj} = 0.55; SAR: pseudo-R^2 = 0.73). Parameter estimates and AICc weights as in Table S4.1.

	OLS				SAR			
Variable	Estimate (standard error)	*z value*	*p*	AICc w	Estimate (standard error)	*z value*	*p*	AICc w
intercept	1115.3 (64.0)	17.2	<0.0001	-	1117.2 (87.2)	12.8	<0.0001	-
AET	-147.6 (64.6)	2.3	0.02	0.88	-166.5 (59.9)	2.8	<0.01	1
Aquifer area	-56.1 (64.2)	0.9	0.86	0.32	-45.1 (53.6)	0.8	0.40	0.29
Log (Elevation range)	-157.1 (70.0)	2.2	0.03	0.88	-192.6 (60.5)	3.2	<0.01	1
Log (Climatic rarity)	152.3 (81.2)	1.9	0.06	0.65	195.1 (69.8)	2.8	<0.01	1
Log (Temperature anomaly)	522.2 (84.9)	6.1	<0.0001	1	590.7 (96.5)	6.1	<0.0001	1
Precipitation anomaly	-73.0 (82.6)	0.9	0.38	0.28	-94.8 (86.0)	1.1	0.27	0.35
Log (Precipitation seasonality)	-70.4 (46.4)	1.5	0.14	0.46	56.1 (56.7)	1	0.32	0.32

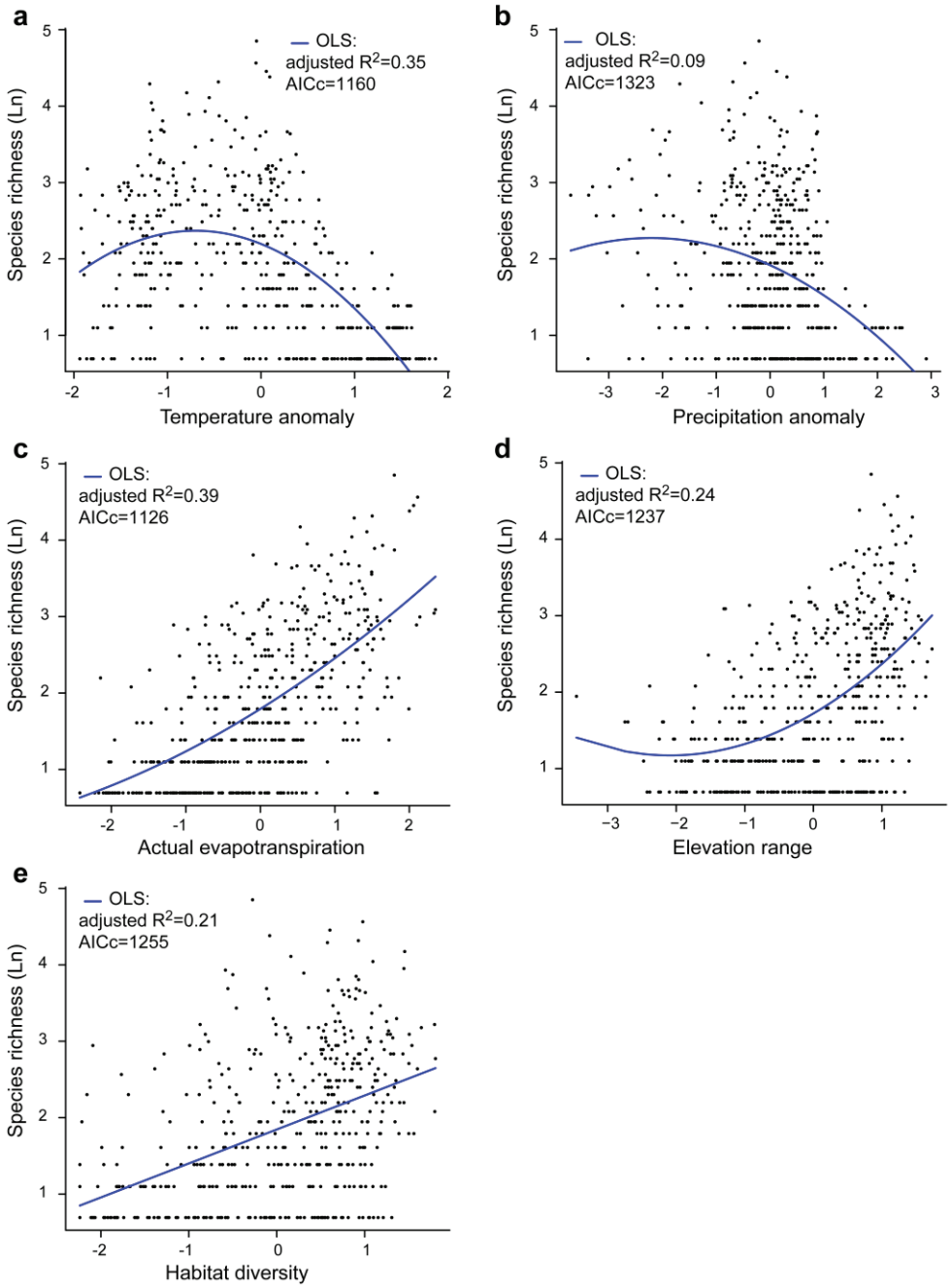
Annexe 4 : Matériel Supplémentaire Article 3

Supplementary material Appendix 1

Eme, D. et al. 2014. Multi-causality and spatial non-stationarity in the determinants of groundwater crustacean diversity in Europe. Submitted to Ecography.
Table A1: Parameter estimates of ordinary least square (OLS) models and simultaneous autoregressive (SAR) models for testing separately the three broad mechanisms on the drivers of species richness of obligate groundwater crustaceans in Europe. Parameter estimates and their standard error are from model averaging using a set of models whose difference in the Akaike's information criterion corrected for small sample size (AICc) with the best model was < 5. AICc*w*: summed AICc weights. For simultaneous autoregressive models, the best neighborhood distance was 300 km.

			Ordinary least square models				Simultaneous autoregressive models			
Mechanisms	Predictors	Terms	AICc*w*	Parameters estimates (SE)	Z value	P value	AICc*w*	Parameters estimates (SE)	Z value	P value
Historical climate stability		Intercept	-	2.21 (0.05)	41.46	<0.0001	-	2.10 (0.20)	10.62	<0.0001
	Temperature anomaly	Linear	1	-0.45 (0.05)	1.96	0.0501	0.68	-0.29 (0.13)	1.69	0.0905
		Quadratic	1	-0.38 (0.04)	9.02	<0.0001	1	-0.41 (0.11)	3.78	<0.0002
	Precipitation anomaly	Linear	0.71	-0.09 (0.05)	8.60	<0.0001	0.59	-0.16 (0.09)	2.27	0.0235
		Quadratic	0.43	0.03 (0.02)	1.19	0.2324	1	0.12 (0.03)	3.80	<0.0002
Productive energy		Intercept	-	1.81 (0.05)	37.24	<0.0001	-	1.82 (0.13)	13.80	<0.0001
	Actual Evapotranspiration	Linear	1	0.61 (0.03)	17.77	<0.0001	1	0.46 (0.06)	7.894	<0.0001
		Quadratic	0.67	0.05 (0.03)	1.86	0.0634	0.32	0.02 (0.04)	0.602	0.5470
Spatial heterogeneity		Intercept	-	1.74 (0.05)	33.68	<0.0001	-	1.81 (0.16)	11.18	<0.0001
	Elevation range	Linear	1	0.41 (0.04)	10.00	<0.0001	1	0.25 (0.05)	5.03	<0.0001
		Quadratic	1	0.10 (0.03)	3.17	0.00154	0.33	0.02 (0.03)	3.24	0.0012
	Habitat diversity	Linear	1	0.33 (0.04)	8.56	<0.0001	1	0.14 (0.04)	1.35	0.1771
		Quadratic	0.26	-8.1e-5 (0.04)	0.002	0.9900	0.47	0.04 (0.03)	0.80	0.4237

Figure A1: Pairwise relationships between species richness (ln transformed) and a) temperature anomaly, b) precipitation anomaly, c) actual evapotranspiration, d) elevation range, and e) habitat diversity. Scale for environmental variable corresponds to standardized values. OLS: ordinary least squares. AICc: Akaike's information criterion corrected for small sample size.



Supplementary material Appendix 2

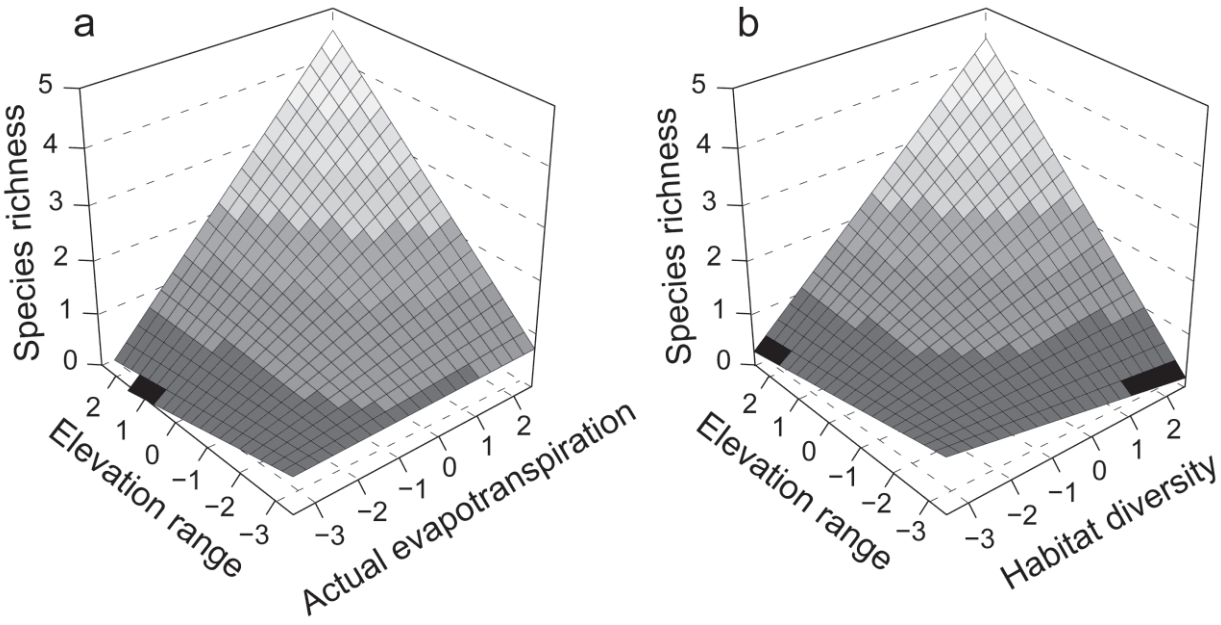
Eme, D. et al. 2014. Multi-causality and spatial non-stationarity in the determinants of groundwater crustacean diversity in Europe. Submitted to Ecography.

Table A2: Parameter estimates of ordinary least square (OLS) models and simultaneous autoregressive (SAR) models for testing jointly the three broad mechanisms on the drivers of species richness of obligate groundwater crustaceans in Europe. Parameter estimates and their standard error (in parenthesis) are from model averaging using a set of models whose difference in the Akaike's information criterion (AICc corrected for small sample size) with the best model was < 5. Ano_T: temperature anomaly; Ano_P: precipitation anomaly; AET: actual evapotranspiration; Elevr: elevation range; Hab: habitat diversity. The superscript next to the predictor's names indicates the quadratic form. For simultaneous autoregressive models, the best neighborhood distance was 220 km.

Model		Intercept	Historical climate stability				Productive energy		Spatial heterogeneity			Interactions
			Ano_T	Ano_T^2	Ano_P	Ano_P^2	AET	AET^2	Elevr	Elevr^2	Hab	
Additive	OLS	1.78 (0.07)	-0.10 (0.05)	-0.07 (0.04)	0.04 (0.04)	- [a]	0.41 (0.04)	0.06 (0.03)	0.25 (0.05)	0.07 (0.03)	0.18 (0.04)	na [b]
	SAR	1.90 (0.11)	-0.11 (0.08)	-0.13 (0.08)	-0.06 (0.07)	0.04 (0.03)	0.38 (0.06)	-	0.20 (0.05)	-	0.15 (0.04)	na
With interactions	OLS	1.85 (0.06)	-0.17 (0.05)	-0.12 (0.04)	-	-	0.32 (0.04)	0.05 (0.03)	0.22 (0.04)	0.01 (0.03)	0.18 (0.04)	Elevr×AET =0.17 (0.04); Elevr×Shannon = 0.12 (0.04)
	SAR	1.91 (0.11)	-0.17 (0.08)	-0.17 (0.08)	-0.10 (0.08)	0.03 (0.03)	0.28 (0.06)	-	0.21 (0.05)	-	0.14 (0.04)	Elevr×AET= 0.18 (0.05); Elevr×Shannon_HA = 0.05 (0.04)

[a] not included in the model selection. [b] na: not applicable

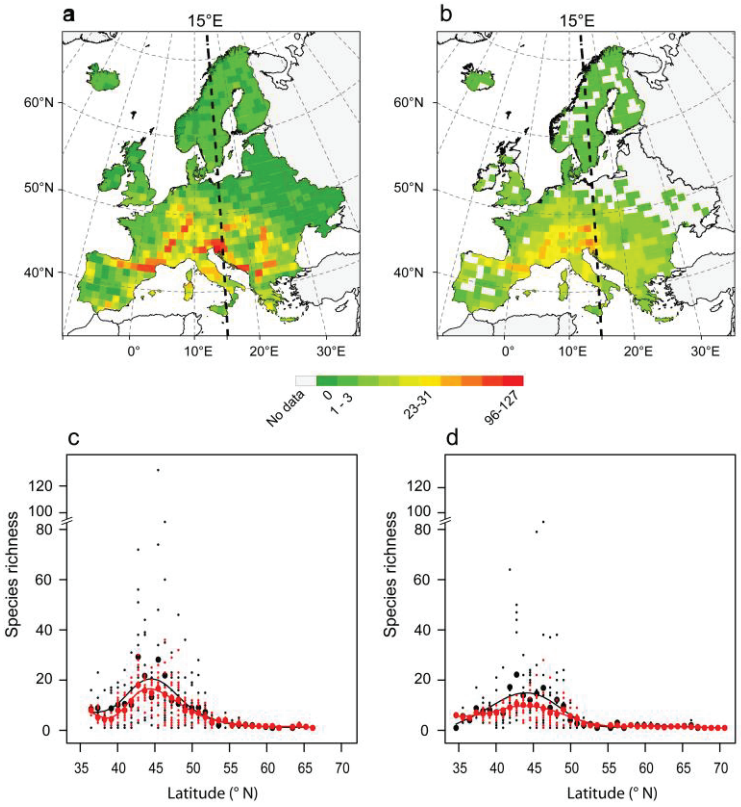
Figure A2: Interaction effects between elevation range (ln transformed) and (a) actual evapotranspiration and (b) habitat diversity on species richness of groundwater crustaceans. Species richness values are fitted values from ordinary least square models. Scale for environmental predictors corresponds to standardized values.



Supplementary material Appendix 3

Eme, D. et al. 2014. Multi-causality and spatial non-stationarity in the determinants of groundwater crustacean diversity in Europe. Submitted to Ecography.

Figure A3: Top row: maps of observed (a) and estimated richness (b). Bottom row: relationships between observed and estimated species richness and latitude for western (< 15° E) (c) and eastern (>15° E) (d) Europe (see broken lines in panels a and b). Estimated richness is from geographically weighted regression. Small black and red dots are observed and estimated species richness per cell, respectively. Large black and red dots are mean observed and estimated species richness per latitudinal band, respectively. Black and red lines in panels represent the fit of a generalized additive model to the cell average per latitudinal band of observed and estimated species richness, respectively.



Annexe 5 : Matériel supplémentaire Article 4

SUPPORTING INFORMATION

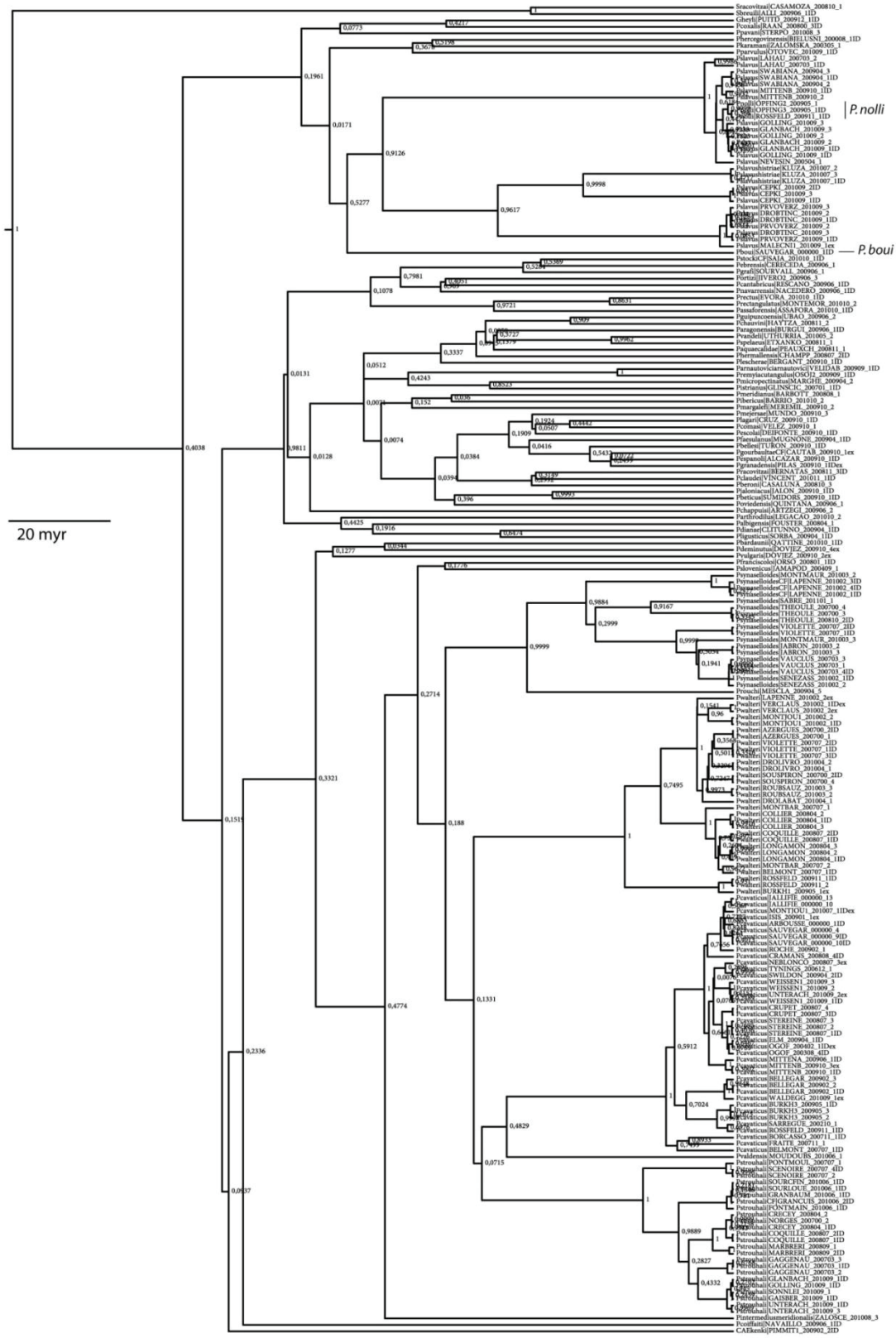
Bayesian phylogeographic inferences reveal contrasting colonization dynamics among European groundwater isopods

D. Eme, F. Malard, L. Konecny-Dupré, T. Lefébure and C. J. Douady

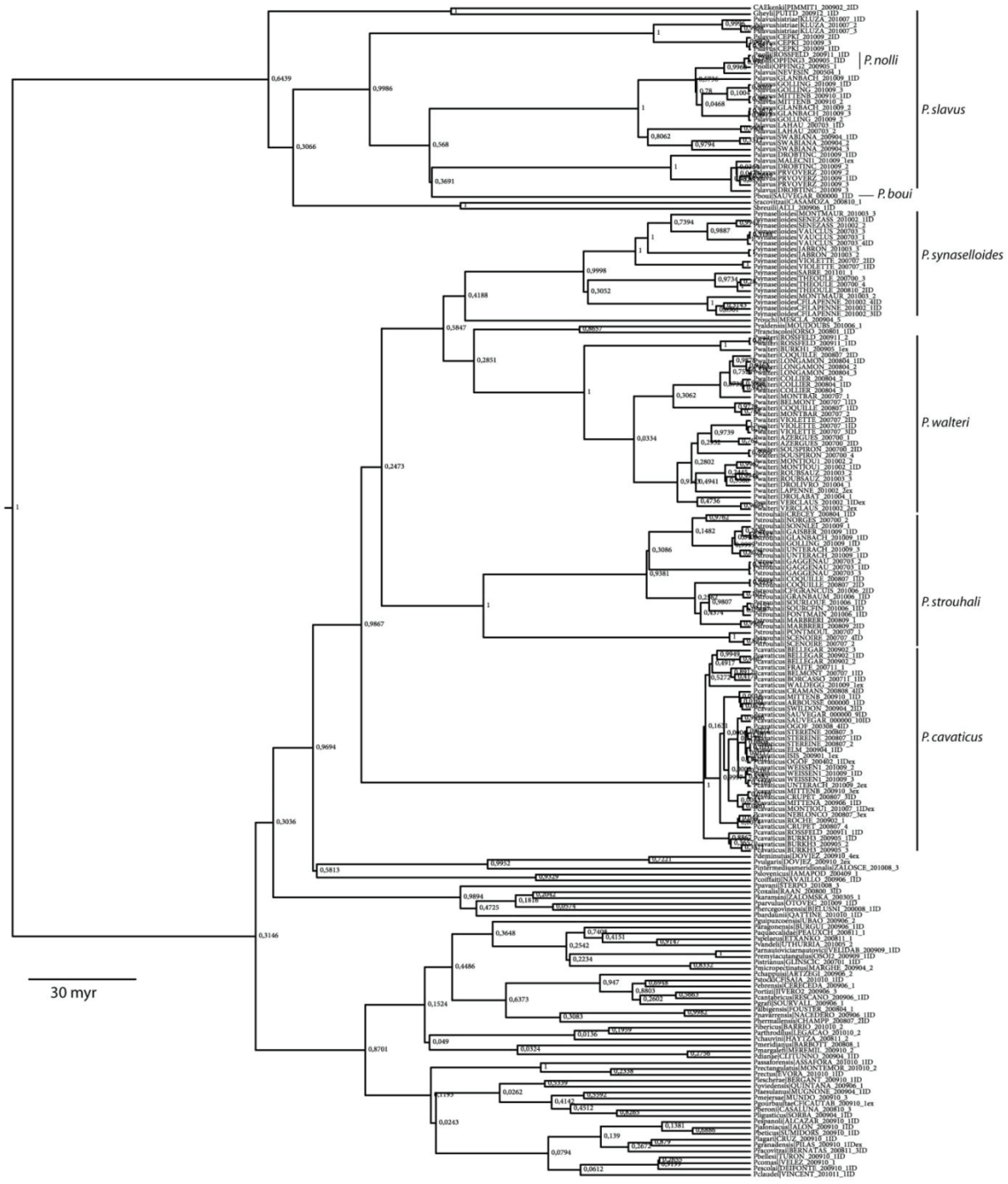
Molecular Ecology

Table of content:

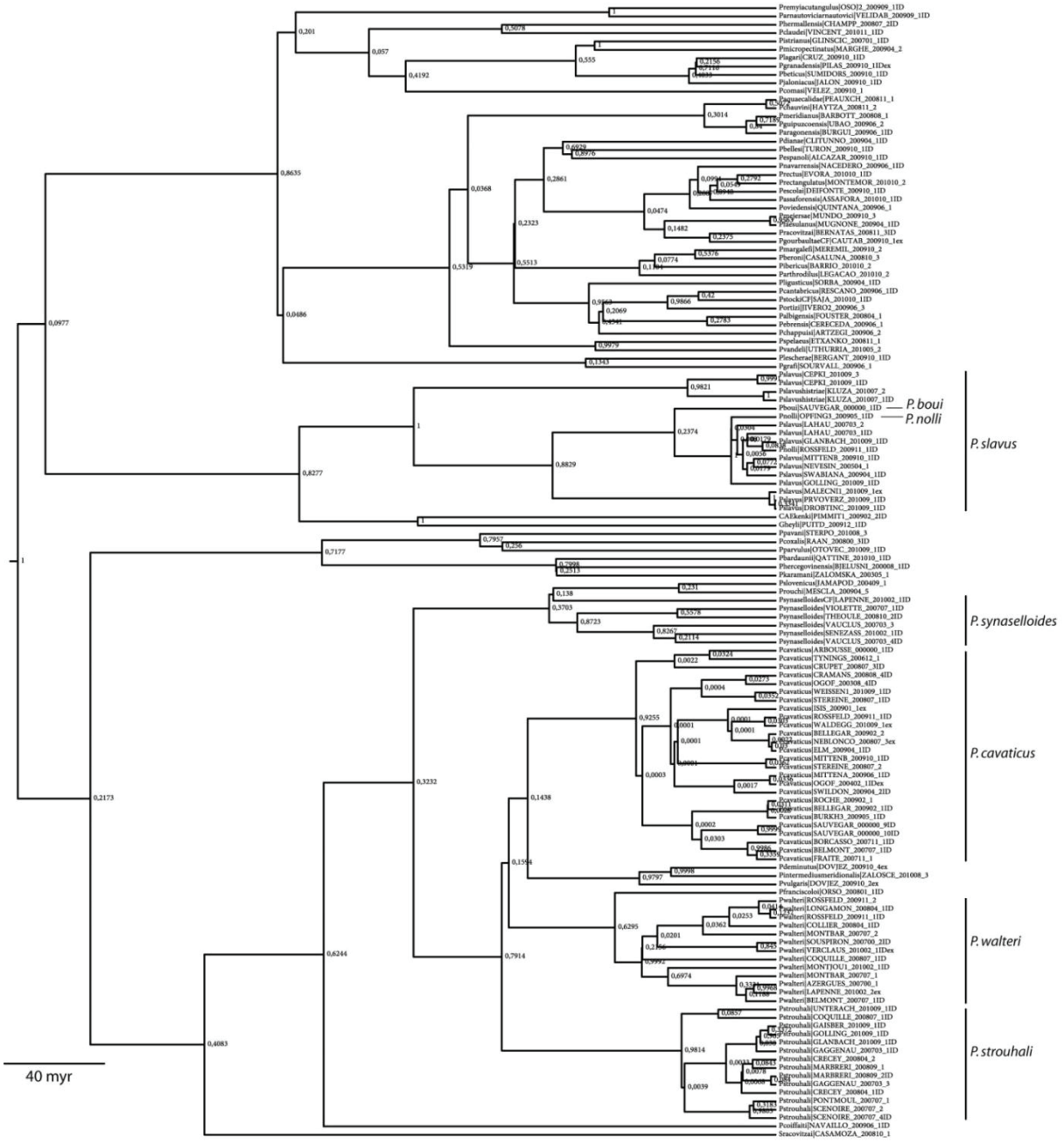
Figure S1 BEAST maximum clade credibility (MCC) gene trees with posterior probability branching support.



a) COI gene tree

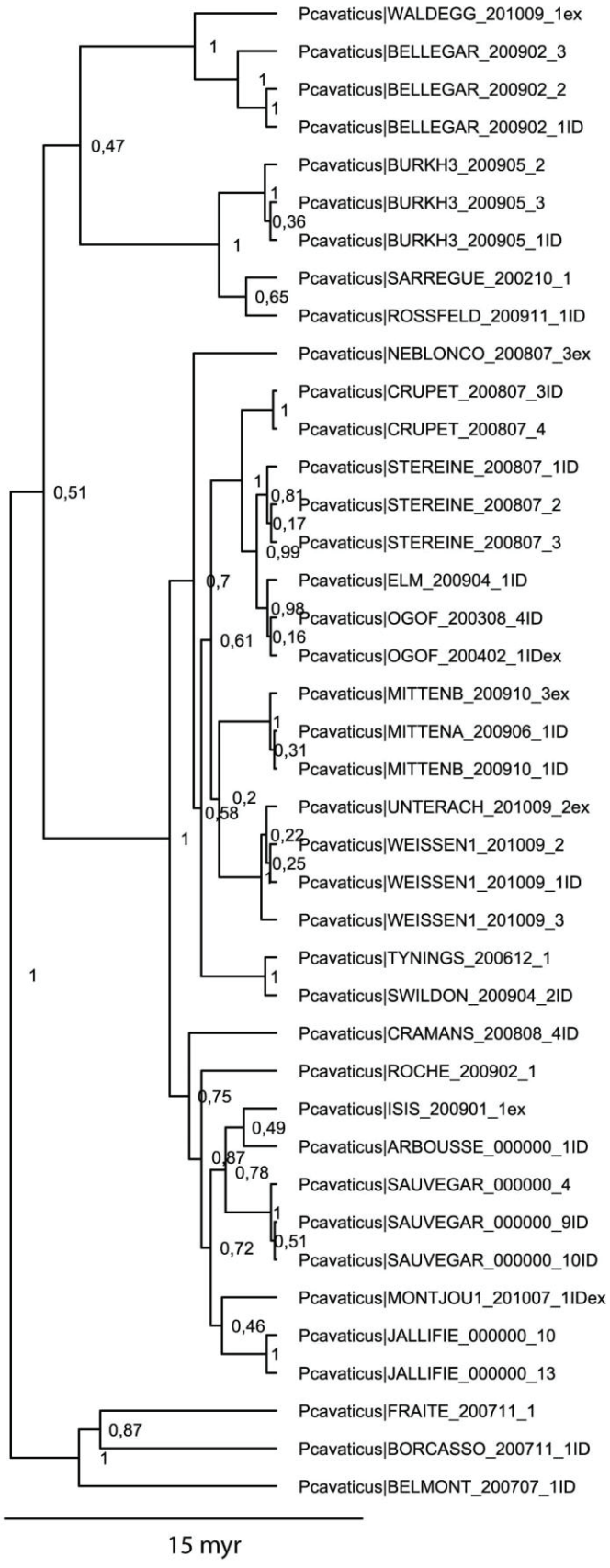


b) 16S gene tree

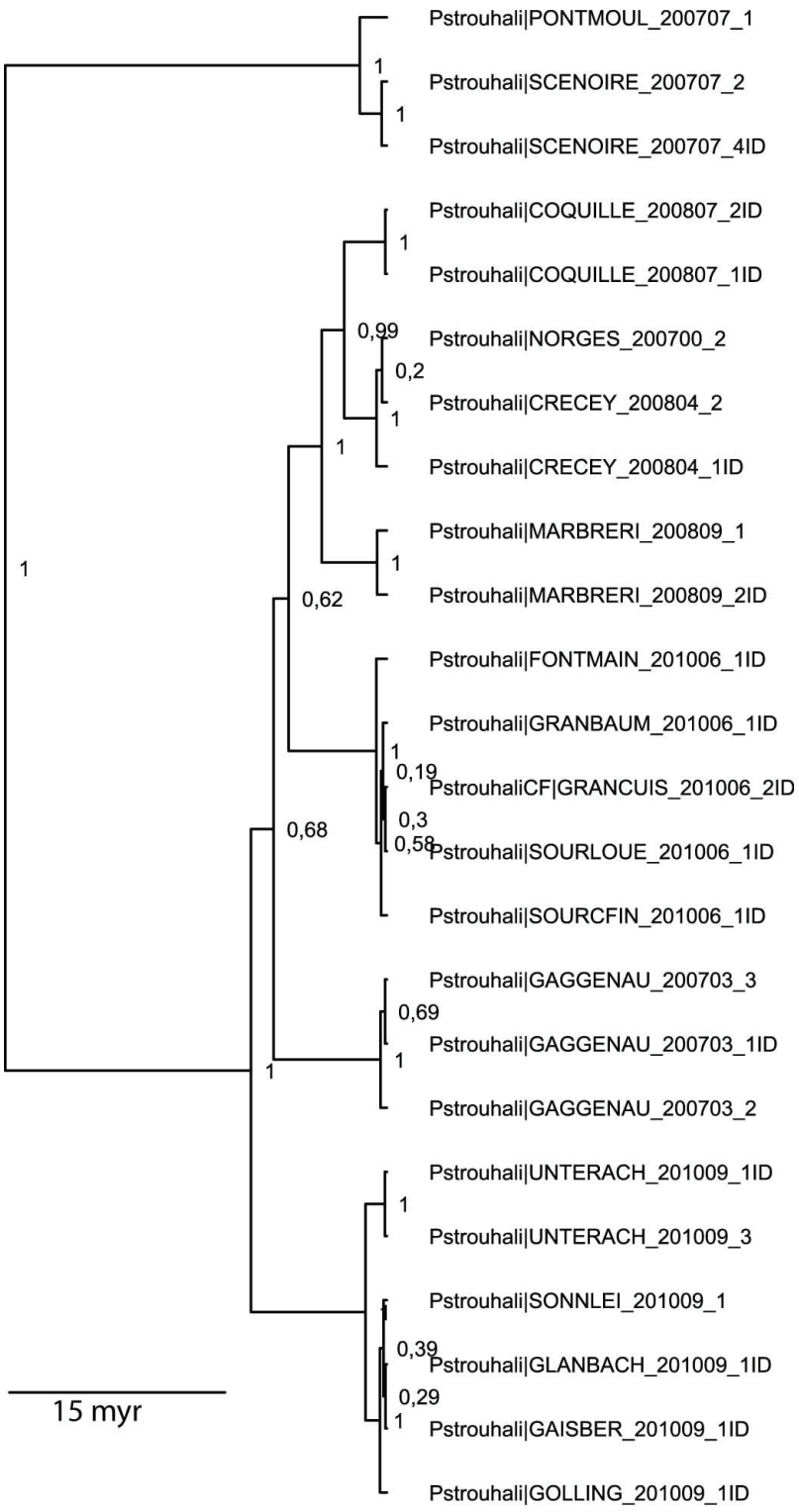


c) 28S gene tree

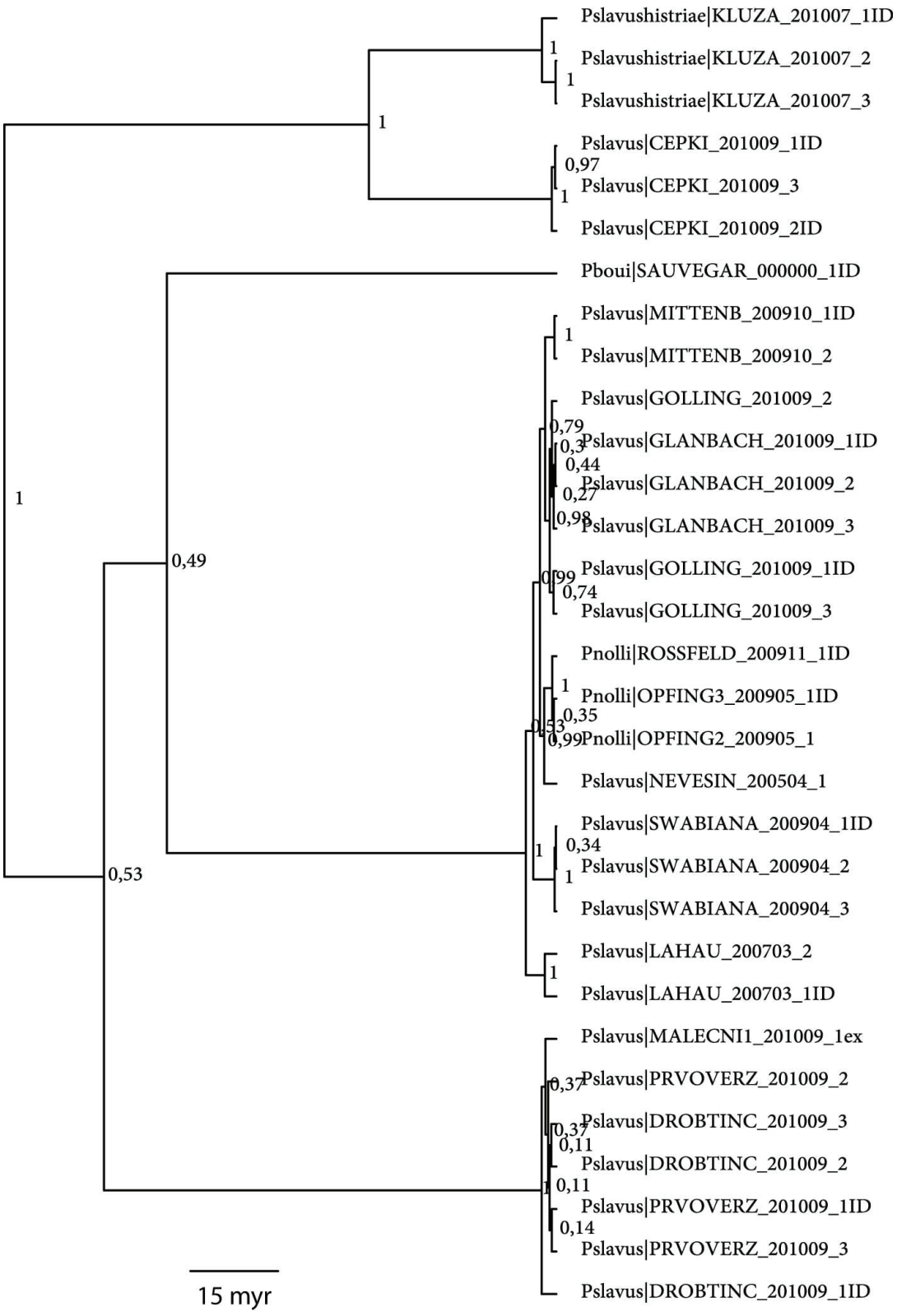
Figure S2 Subsets of the maximum clade credibility tree with linked topology showing branching patterns within focal taxa.



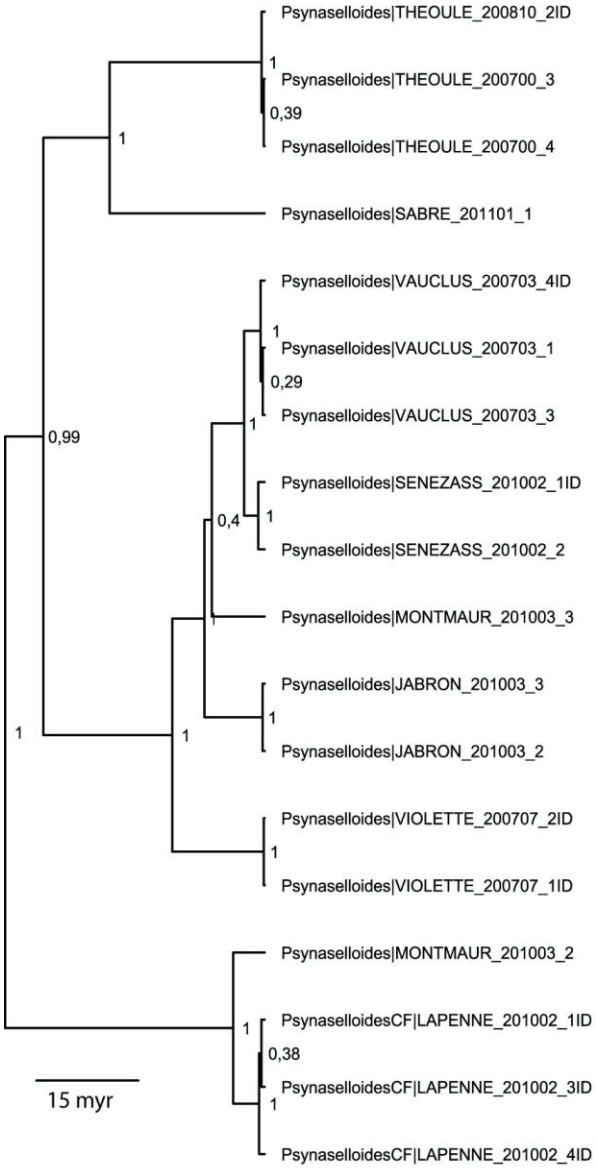
a) *P. cavaticus*



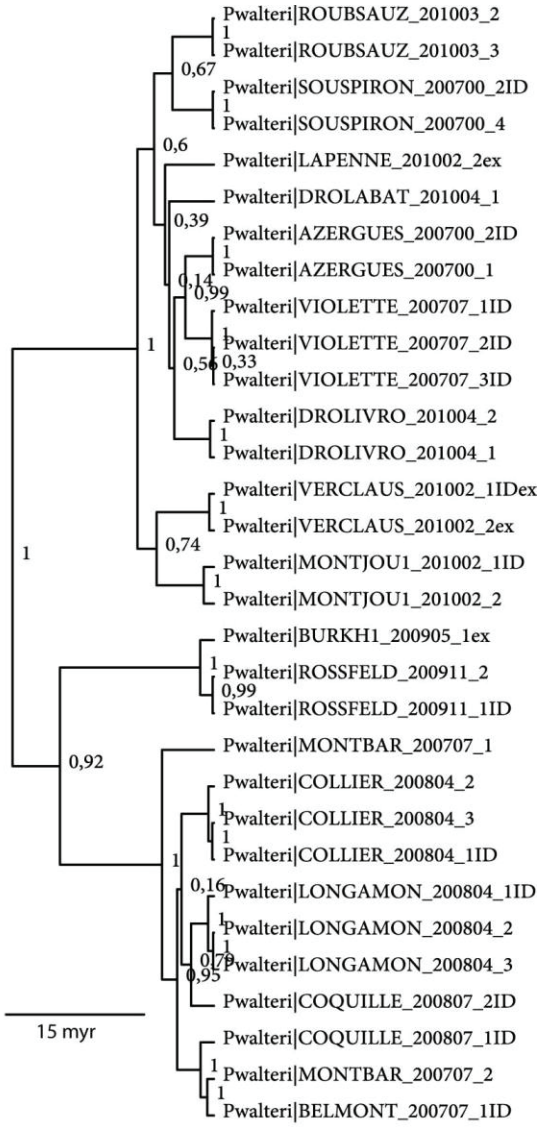
b) *P. strouhali*



c) *P. slavus* morphospecies including *P. nolli* and *P. boui*



d) *P. synaselloides*



e) *P. walteri*

Figure S3 28S genes trees of the 5 focal morphospecies reconstruct with PhyML and GTR+G+I evolution model showing the distribution of *28S* haplotypes among *COI* cryptic species.

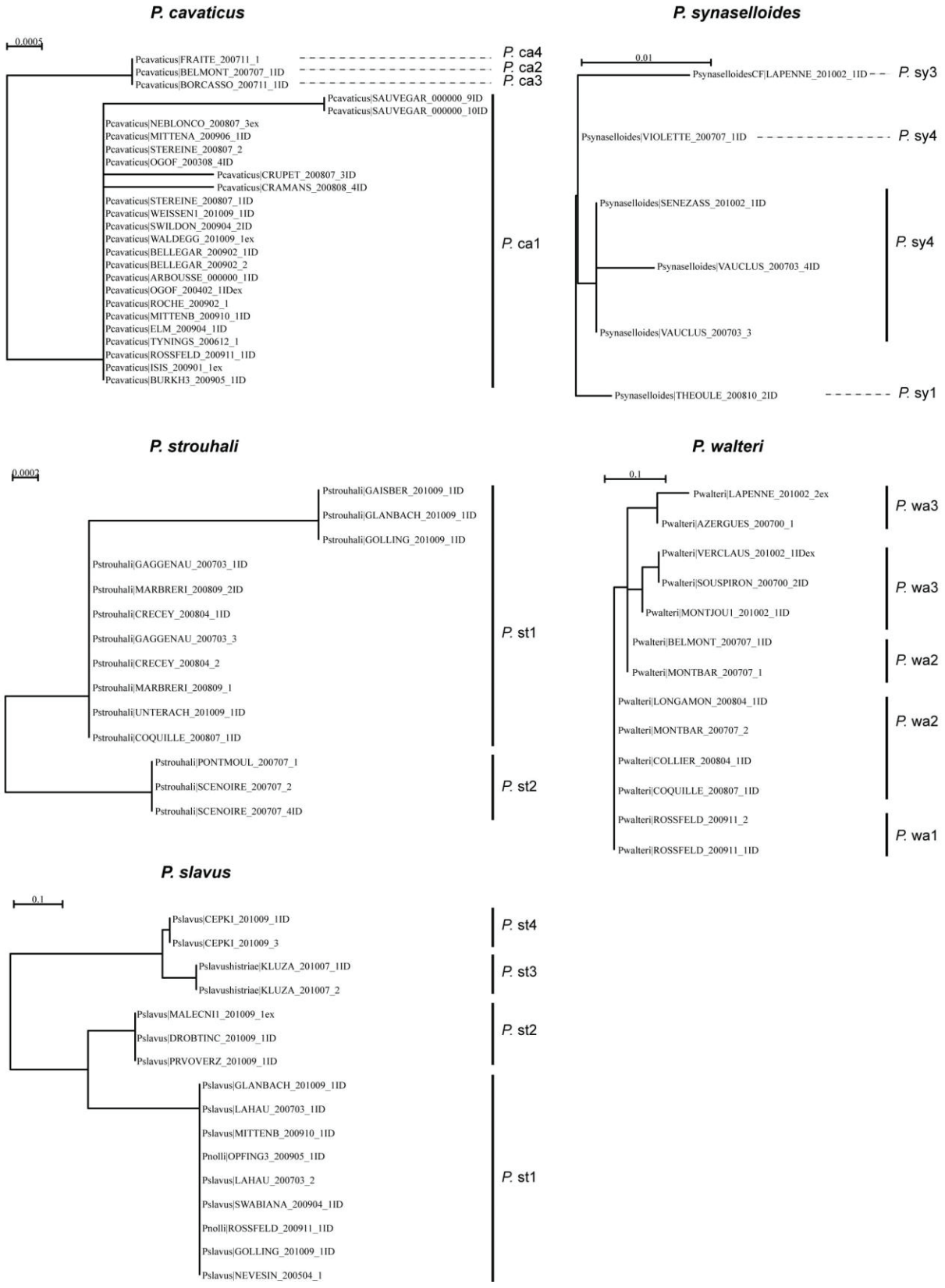
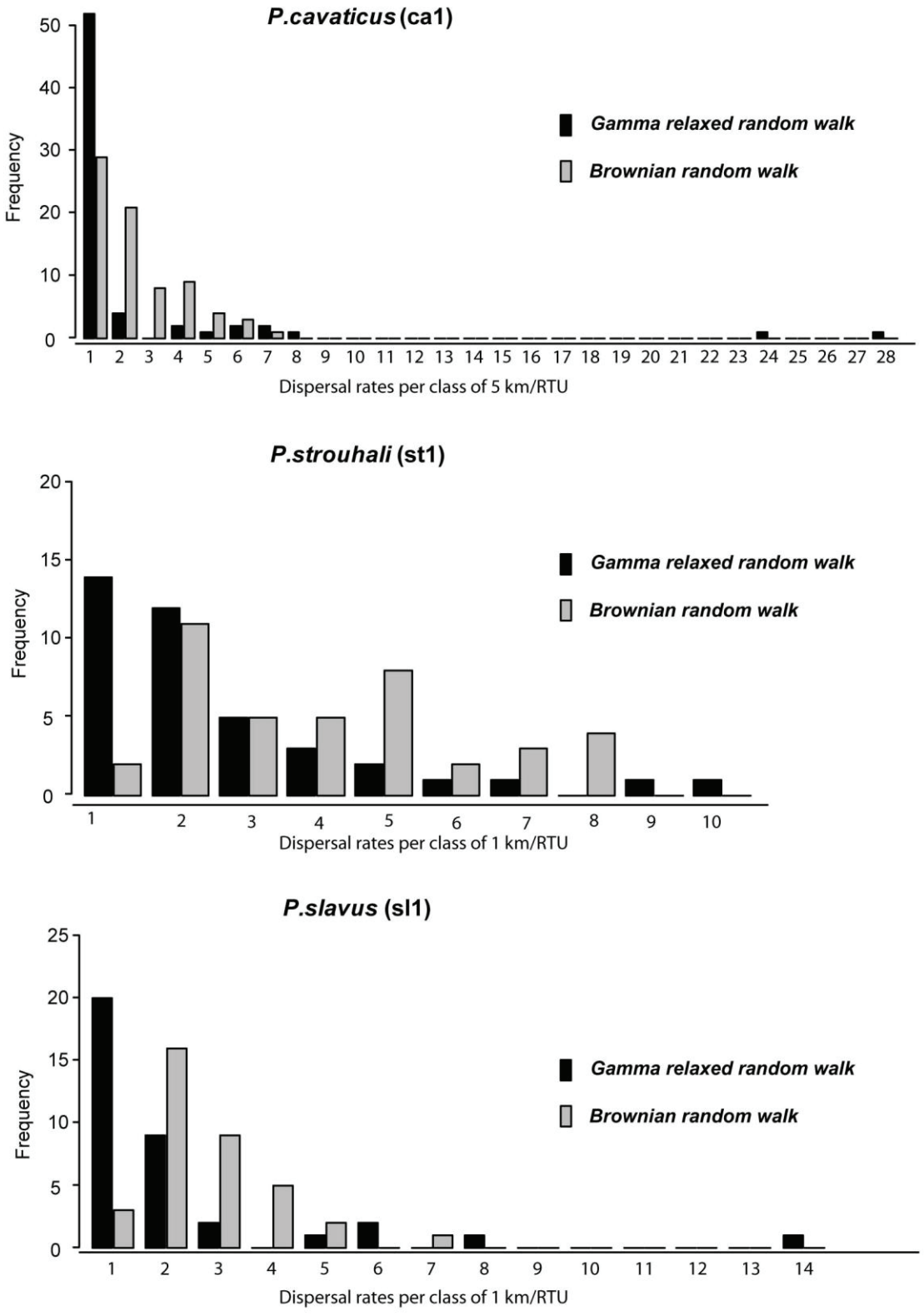


Figure S4 Frequency distribution of dispersal rates inferred by a Brownian random walk model and a Gamma relaxed random walk model for the 3 widely-distributed species. Dispersal is measured as relative time unit (RTU) using the MRCA of *P. cavaticus* (ca1) as a reference. Compared to a Brownian distribution, using a Gamma distribution allows larger variation of dispersal rates for the three species.



SUPPORTING INFORMATION

Bayesian phylogeographic inferences reveal contrasting colonization dynamics among European groundwater isopods

D. Eme, F. Malard, L. Konecny-Dupré, T. Lefébure and C. J. Douady

Molecular Ecology

Table of content:

Table S1 Sequence data set. CO : country name, LA: latitude in decimal degrees (N), LO: longitude in decimal degrees (E), HA: code of haplotypes used for testing the monophyly, HB: code CO1 haplotype used for species delineation , CO1_TH: code of cryptic species delimited by the CO1 threshold, CO1_GM: code of cryptic species delimited by the GMYC , ANCO1: accession number for CO1 sequence, AN16S: accession number for 16S sequence, AN28S: accession number for 28S sequence. In bold, sequences coming from this study.

SPECIES	SUB-SPECIES	INDIVIDUAL	LOCALITY	CO	LA	LO	HA	HB	CO1_TH	CO1_GM	ANCO1	AN16S	AN28S
Ingroup													
Proasellus cavaticus (Leydig, 1871)	no	ARBOUSSE_000000_1ID	Ceilhes et rocozels	FRANCE	43.82250	3.14361	Ha_13	Hb_13	ca1	ca1g	JQ921087†	**KC610160**	**KC610369**
	no	BELLEGAR_200902_1ID	Jaun	SWITZERLAND	46.61139	7.27583	Ha_14	Hb_14	ca1	ca1g	JQ921073†	**KC610161**	**KC610370**
	no	BELLEGAR_200902_2	Jaun	SWITZERLAND	46.61139	7.27583	Ha_15	Hb_15	ca1	ca1g	JQ921074†	**KC610162**	**KC610371**
	no	BELLEGAR_200902_3	Jaun	SWITZERLAND	46.61139	7.27583	Ha_16	Hb_16	ca1	ca1g	JQ921075†	**KC610163**	x
	no	BELMONT_200707_1ID	Belmont	FRANCE	47.00945	5.59472	Ha_17	Hb_17	ca2	ca2g	JQ921076†	JQ921722†	JQ921904†
	no	BORCASSO_200711_1ID	Nevy sur seille	FRANCE	46.73667	5.64222	Ha_18	Hb_18	ca3	ca2g	JQ921077†	JQ921723†	JQ921905†
	no	BURKH3_200905_1ID	Burkheim	GERMANY	48.09513	7.59328	Ha_19	Hb_19	ca1	ca1g	KC610465§	KC610093§	KC610340§
	no	BURKH3_200905_2	Burkheim	GERMANY	48.09513	7.59328	Ha_20	Hb_19	ca1	ca1g	JQ921078†	**KC610164**	x
	no	BURKH3_200905_3	Burkheim	GERMANY	48.09513	7.59328	Ha_21	Hb_19	ca1	ca1g	JQ921079†	**KC610165**	x
	no	CRAMANS_200808_4ID	Challes la montagne	FRANCE	46.13556	5.47611	Ha_22	Hb_20	ca1	ca1g	JQ921088†	**KC610166**	**KC610372**
	no	CRUPET_200807_3ID	Assesse	BELGIUM	50.35449	4.96145	Ha_23	Hb_21	ca1	ca1g	JQ921080†	**KC610167**	**KC610373**
	no	CRUPET_200807_4	Assesse	BELGIUM	50.35449	4.96145	Ha_24	Hb_21	ca1	ca1g	JQ921081†	**KC610168**	x
	no	ELM_200904_1ID	Clydach	UNITED KINGDOM	51.80521	-3.13996	Ha_25*	Hb_22	ca1	ca1g	JQ921082†	**KC610169**	**KC610374**
	no	FRAITE_200711_1	Thoiria	FRANCE	46.52500	5.73389	Ha_26	Hb_23	ca4	ca2g	JQ921072†	JQ921721†	JQ921903†
	no	ISIS_200901_1ex	Aveze	FRANCE	43.98361	3.59806	Ha_27	Hb_24	ca1	ca1g	JQ921083†	**KC610170**	**KC610375**
	no	JALLIFIE_000000_10	Choranche	FRANCE	45.07500	5.38528	Ha_28	Hb_25	ca1	ca1g	JQ921084†	x	x
	no	JALLIFIE_000000_13	Choranche	FRANCE	45.07500	5.38528	Ha_29	Hb_26	ca1	ca1g	JQ921085†	x	x
	no	MITTENA_200906_1ID	Mittenwald	GERMANY	47.42670	11.25920	Ha_30	Hb_27	ca1	ca1g	JQ921089†	**KC610171**	**KC610376**
	no	MITTENB_200910_1ID	Mittenwald	GERMANY	47.42670	11.25920	Ha_31	Hb_28	ca1	ca1g	JQ921090†	**KC610172**	**KC610377**
	no	MITTENB_200910_3ex	Mittenwald	GERMANY	47.42670	11.25920	Ha_32	Hb_27	ca1	ca1g	JQ921091†	**KC610173**	x
	no	MONTJOU1_201007_1IDex	Montjoux	FRANCE	44.50177	5.08935	Ha_33	Hb_29	ca1	ca1g	**KC610479**	**KC610174**	x
	no	NEBLONCO_200807_3ex	Jenneret	BELGIUM	50.41377	5.45694	Ha_34	Hb_30	ca1	ca1g	JQ921092†	**KC610175**	**KC610378**
	no	OGOF_200308_38ml	Glyntawe	UNITED KINGDOM	51.82441	-3.67296	Ha_25	Hb_22	ca1	ca1g	**KC610480**	x	x
	no	OGOF_200308_4ID	Glyntawe	UNITED KINGDOM	51.82441	-3.67296	Ha_36	Hb_31	ca1	ca1g	JQ921093†	**KC610176**	**KC610379**
	no	OGOF_200402_1IDex	Glyntawe	UNITED KINGDOM	51.82441	-3.67296	Ha_37	Hb_22	ca1	ca1g	JQ921100†	JQ921725†	JQ921907†
	no	OGOF_200308_2IDex	Glyntawe	UNITED KINGDOM	51.82441	-3.67296	Ha_25	Hb_22	ca1	ca1g	**KC610481**	x	x
	no	ROCHE_200902_1	Rencruel	FRANCE	45.07167	5.49750	Ha_39	Hb_32	ca1	ca1g	JQ921097†	**KC610177**	**KC610380**
	no	ROSSFELD_200911_1ID	Rossfeld	FRANCE	48.33390	7.63116	Ha_40	Hb_33	ca1	ca1g	JQ921086†	JQ921724†	JQ921906†
	no	SARREGUE_200210_1	Sarreguemines	FRANCE	49.11516	7.03750	Ha_41	Hb_34	ca1	ca1g	**KC610482**	x	x
	no	SAUVEGAR_000000_10ID	Sauve	FRANCE	43.94056	3.95000	Ha_42*	Hb_35	ca1	ca1g	JQ921099†	**KC610178**	**KC610381**
	no	SAUVEGAR_000000_4	Sauve	FRANCE	43.94056	3.95000	Ha_43	Hb_35	ca1	ca1g	**KC610483**	x	x
	no	SAUVEGAR_000000_5	Sauve	FRANCE	43.94056	3.95000	Ha_42	Hb_35	ca1	ca1g	**KC610484**	x	x
	no	SAUVEGAR_000000_9ID	Sauve	FRANCE	43.94056	3.95000	Ha_45	Hb_35	ca1	ca1g	JQ921098†	**KC610179**	**KC610382**
	no	STEREINE_200807_1ID	Pierre la treiche	FRANCE	48.64528	5.94194	Ha_46	Hb_36	ca1	ca1g	JQ921094†	**KC610180**	**KC610383**
	no	STEREINE_200807_2	Pierre la treiche	FRANCE	48.64528	5.94194	Ha_47	Hb_37	ca1	ca1g	JQ921095†	**KC610181**	**KC610384**
	no	STEREINE_200807_3	Pierre la treiche	FRANCE	48.64528	5.94194	Ha_48	Hb_38	ca1	ca1g	JQ921096†	**KC610182**	x
	no	SWILDON_200904_2ID	Priddy	UNITED KINGDOM	51.25896	-2.67324	Ha_49	Hb_39	ca1	ca1g	JQ921101†	**KC610183**	**KC610385**
	no	TYNINGS_200612_1	Tyning	UNITED KINGDOM	51.30407	-2.75980	Ha_50	Hb_40	ca1	ca1g	JQ921102†	x	**KC610386**
	no	UNTERACH_201009_2ex	Mondsee	AUSTRIA	47.79230	13.48452	Ha_51	Hb_41	ca1	ca1g	JQ921103†	**KC610184**	x
	no	WALDEGG_201009_1ex	Interlaken	SWITZERLAND	46.71456	7.83211	Ha_52	Hb_42	ca1	ca1g	JQ921104†	JQ921726†	JQ921908†
	no	WEISSEN1_201009_1ID	Mondsee	AUSTRIA	47.78919	13.55226	Ha_53	Hb_43	ca1	ca1g	JQ921105†	**KC610185**	**KC610387**
	no	WEISSEN1_201009_2	Mondsee	AUSTRIA	47.78919	13.55226	Ha_54	Hb_44	ca1	ca1g	JQ921106†	**KC610186**	x
	no	WEISSEN1_201009_3	Mondsee	AUSTRIA	47.78919	13.55226	Ha_55	Hb_45	ca1	ca1g	JQ921107†	**KC610187**	x
Proasellus nolli (Karaman, 1952)	no	BURKH1_200905_1	Burkheim	GERMANY	48.09162	7.59935	Ha_90	Hb_78	sl1	sl1g	JQ921331†	x	x
	no	OPFING1_200905_1	Opfingen	GERMANY	48.00093	7.75635	Ha_90	Hb_78	sl1	sl1g	JQ921333†	**KC610188**	x
	no	OPFING2_200905_1	Opfingen	GERMANY	48.00713	7.76364	Ha_90*	Hb_78	sl1	sl1g	JQ921334†	**KC610189**	x
	no	OPFING3_200905_1ID	Opfingen	GERMANY	48.00946	7.76707	Ha_91	Hb_79	sl1	sl1g	JQ921335†	**KC610190**	**KC610388**
	no	ROSSFELD_200911_1ID	Rossfeld	FRANCE	48.33390	7.63116	Ha_92	Hb_80	sl1	sl1g	JQ921332†	JQ921784†	JQ921965†
Proasellus slavus (Remy, 1948)	*slavus histriae* (Sket, 1963)	CEPKI_201009_1ID	Cepki	SLOVENIA	45.54328	13.83449	Ha_102	Hb_90	sl4	sl4g	JQ921386†	JQ921806†	JQ921986†
	slavus histriae (Sket, 1963)	CEPKI_201009_2ID	Cepki	SLOVENIA	45.54328	13.83449	Ha_103	Hb_91	sl4	sl4g	JQ921387†	**KC610191**	x
	slavus histriae (Sket, 1963)	CEPKI_201009_3	Cepki	SLOVENIA	45.54328	13.83449	Ha_104	Hb_91	sl4	sl4g	JQ921388†	**KC610192**	**KC610389**
	slavus ssp. "Mura" (Sket, 1963)	DROBTINC_201009_1ID	Apace	SLOVENIA	46.69689	15.84603	Ha_105	Hb_92	sl2	sl2g	JQ921389†	**KC610193**	**KC610390**
	slavus ssp. "Mura" (Sket, 1963)	DROBTINC_201009_2	Apace	SLOVENIA	46.69689	15.84603	Ha_106	Hb_93	sl2	sl2g	JQ921390†	**KC610194**	x

	slavus ssp. "Mura" (Sket, 1963)	DROBTINC_201009_3	Apace	SLOVENIA	46.69689	15.84603	Ha_107	Hb_94	sl2	sl2g	JQ921391†	**KC610195**	x
	slavus salisburgiensis (Strouhal, 1958)	GLANBACH_201009_1ID	Salzburg	AUSTRIA	47.81010	13.02193	Ha_108	Hb_95	sl1	sl1g	JQ921400†	JQ921807†	JQ921987†
	slavus salisburgiensis (Strouhal, 1958)	GLANBACH_201009_2	Salzburg	AUSTRIA	47.81010	13.02193	Ha_109	Hb_95	sl1	sl1g	JQ921401†	**KC610196**	x
	slavus salisburgiensis (Strouhal, 1958)	GLANBACH_201009_3	Salzburg	AUSTRIA	47.81010	13.02193	Ha_110	Hb_96	sl1	sl1g	JQ921402†	**KC610197**	x
	slavus salisburgiensis (Strouhal, 1958)	GOLLING_201009_1ID	Salzburg	AUSTRIA	47.58191	13.17079	Ha_111	Hb_95	sl1	sl1g	JQ921403†	**KC610198**	**KC610391**
	slavus salisburgiensis (Strouhal, 1958)	GOLLING_201009_2	Salzburg	AUSTRIA	47.58191	13.17079	Ha_112	Hb_97	sl1	sl1g	JQ921404†	**KC610199**	x
	slavus salisburgiensis (Strouhal, 1958)	GOLLING_201009_3	Salzburg	AUSTRIA	47.58191	13.17079	Ha_113	Hb_98	sl1	sl1g	JQ921405†	**KC610200**	x
	slavus histriae (Sket, 1963)	KLUZA_201007_1ID	Trieste	ITALY	45.60010	13.85440	Ha_126	Hb_108	sl3	sl3g	JQ921410†	**KC610212**	**KC610398**
	slavus histriae (Sket, 1963)	KLUZA_201007_2	Trieste	ITALY	45.60010	13.85440	Ha_127	Hb_109	sl3	sl3g	JQ921411†	JQ921808†	JQ921988†
	slavus histriae (Sket, 1963)	KLUZA_201007_3	Trieste	ITALY	45.60010	13.85440	Ha_128	Hb_110	sl3	sl3g	JQ921412†	**KC610213**	x
	slavus slavus (Remy, 1948)	LAHAU_200703_1ID	Vienne	AUSTRIA	48.18430	16.52240	Ha_114	Hb_99	sl1	sl1g	JQ921393†	**KC610201**	**KC610392**
	slavus slavus (Remy, 1948)	LAHAU_200703_2	Vienne	AUSTRIA	48.18430	16.52240	Ha_115	Hb_100	sl1	sl1g	JQ921394†	**KC610202**	**KC610393**
	slavus styriacus (Sket, 1963)	MALECN11_201009_1ex	Maribor	SLOVENIA	46.55428	15.69496	Ha_116	Hb_101	sl2	sl2g	JQ921406†	**KC610203**	**KC610394**
	slavus slavus (Remy, 1948)	MITTENB_200910_1ID	Mittenwald	GERMANY	47.42670	11.25920	Ha_117	Hb_102	sl1	sl1g	JQ921395†	**KC610204**	**KC610395**
	slavus slavus (Remy, 1948)	MITTENB_200910_2	Mittenwald	GERMANY	47.42670	11.25920	Ha_118	Hb_103	sl1	sl1g	JQ921396†	**KC610205**	x
	slavus slavus (Remy, 1948)	NEVESIN_200504_1	Landau	GERMANY	49.19361	8.11361	Ha_119	Hb_104	sl1	sl1g	JQ921392†	**KC610206**	**KC610396**
	slavus ssp. "Mura" (Sket, 1963)	PRVOVERZ_201009_1ID	Verzej	SLOVENIA	46.58587	16.16869	Ha_120	Hb_92	sl2	sl2g	JQ921407†	JQ921809†	JQ921989†
	slavus ssp. "Mura" (Sket, 1963)	PRVOVERZ_201009_2	Verzej	SLOVENIA	46.58587	16.16869	Ha_121	Hb_105	sl2	sl2g	JQ921408†	**KC610207**	x
	slavus ssp. "Mura" (Sket, 1963)	PRVOVERZ_201009_3	Verzej	SLOVENIA	46.58587	16.16869	Ha_122	Hb_106	sl2	sl2g	JQ921409†	**KC610208**	x
	slavus slavus (Remy, 1948)	SWABIANA_200904_1ID	Bachingen an der brenz	GERMANY	48.53250	10.28500	Ha_123	Hb_107	sl1	sl1g	JQ921397†	**KC610209**	**KC610397**
	slavus slavus (Remy, 1948)	SWABIANA_200904_2	Bachingen an der brenz	GERMANY	48.53250	10.28500	Ha_124	Hb_107	sl1	sl1g	JQ921398†	**KC610210**	x
	slavus slavus (Remy, 1948)	SWABIANA_200904_3	Bachingen an der brenz	GERMANY	48.53250	10.28500	Ha_125	Hb_107	sl1	sl1g	JQ921399†	**KC610211**	x
Proasellus strouhali (Karaman, 1955)	*strouhali puteanus* (Henry, 1966)	COQUILLE_200807_1ID	Etalante	FRANCE	47.64500	4.76611	Ha_132	Hb_114	st1	st1g	JQ921419†	**KC610214**	**KC610399**
	strouhali puteanus (Henry, 1966)	COQUILLE_200807_2ID	Etalante	FRANCE	47.64500	4.76611	Ha_133	Hb_114	st1	st1g	JQ921420†	**KC610215**	x
	strouhali puteanus (Henry, 1966)	CRECEY_200804_1ID	Crecey sur tille	FRANCE	47.56333	5.12611	Ha_134	Hb_115	st1	st1g	JQ921421†	**KC610216**	**KC610400**
	strouhali puteanus (Henry, 1966)	CRECEY_200804_2	Crecey sur tille	FRANCE	47.56333	5.12611	Ha_135	Hb_115	st1	st1g	JQ921422†	x	**KC610401**
	strouhali puteanus (Henry, 1966)	FONTMAIN_201006_1ID	Scey maisieres	FRANCE	47.09351	6.05768	Ha_136	Hb_116	st1	st1g	**KC610485**	**KC610217**	x
	strouhali puteanus (Henry, 1966)	GAGGENAU_200703_1ID	Gaggenau	GERMANY	48.80280	8.32654	Ha_137	Hb_117	st1	st1g	JQ921423†	**KC610218**	**KC610402**
	strouhali puteanus (Henry, 1966)	GAGGENAU_200703_2	Gaggenau	GERMANY	48.80280	8.32654	Ha_138	Hb_118	st1	st1g	JQ921424†	**KC610219**	x
	strouhali puteanus (Henry, 1966)	GAGGENAU_200703_3	Gaggenau	GERMANY	48.80280	8.32654	Ha_139	Hb_117	st1	st1g	JQ921425†	JQ921813†	JQ921992†
	strouhali strouhali (Karaman, 1955)	GAISBER_201009_1ID	Mondsee	AUSTRIA	47.84830	13.33237	Ha_140	Hb_119	st1	st1g	JQ921428†	**KC610220**	**KC610403**
	strouhali strouhali (Karaman, 1955)	GLANBACH_201009_1ID	Salzburg	AUSTRIA	47.81010	13.02193	Ha_141	Hb_120	st1	st1g	JQ921435†	**KC610221**	**KC610404**
	strouhali strouhali (Karaman, 1955)	GOLLING_201009_1ID	Salzburg	AUSTRIA	47.58191	13.17079	Ha_142	Hb_121	st1	st1g	JQ921436†	JQ921814†	JQ921993†
	strouhali puteanus (Henry, 1966)	GRANBAUM_201006_1ID	Lods	FRANCE	47.04332	6.23755	Ha_143	Hb_122	st1	st1g	JQ921439†	**KC610222**	x
	strouhali puteanus (Henry, 1966)	GRANCUIS_201006_2ID	La chatelaine	FRANCE	46.87901	5.81454	Ha_155	Hb_131	st1	st1g	JQ921119†	JQ921812†	x
	strouhali puteanus (Henry, 1966)	MARBRERI_200809_1	Perreon	FRANCE	46.05750	4.56294	Ha_144	Hb_123	st1	st1g	JQ921426†	**KC610223**	**KC610405**
	strouhali puteanus (Henry, 1966)	MARBRERI_200809_2ID	Perreon	FRANCE	46.05750	4.56294	Ha_145	Hb_124	st1	st1g	JQ921427†	KC610114§	KC610361§
	strouhali puteanus (Henry, 1966)	NORGES_200700_2	Norges la ville	FRANCE	47.40694	5.07861	Ha_146	Hb_125	st1	st1g	JQ921429†	**KC610224**	x
	strouhali puteanus (Henry, 1966)	PONTMOUL_200707_1	Pont les moulins	FRANCE	47.31750	6.36083	Ha_147	Hb_126	st2	st2g	JQ921430†	**KC610225**	**KC610406**
	strouhali puteanus (Henry, 1966)	SCENOIRE_200707_2	Cusance	FRANCE	47.32083	6.43917	Ha_148	Hb_127	st2	st2g	JQ921431†	**KC610226**	**KC610407**
	strouhali puteanus (Henry, 1966)	SCENOIRE_200707_4ID	Cusance	FRANCE	47.32083	6.43917	Ha_149	Hb_128	st2	st2g	JQ921432†	JQ921816†	JQ921995†
	strouhali strouhali (Karaman, 1955)	SONNLEI_201009_1	Mondsee	AUSTRIA	47.84972	13.32094	Ha_150	Hb_129	st1	st1g	**KC610486**	**KC610227**	x
	strouhali puteanus (Henry, 1966)	SOURCFIN_201006_1ID	Mesnay	FRANCE	46.89295	5.80548	Ha_151	Hb_122	st1	st1g	JQ921437†	**KC610228**	x
	strouhali puteanus (Henry, 1966)	SOURLOUE_201006_1ID	Ouhans	FRANCE	47.01108	6.29944	Ha_152	Hb_122	st1	st1g	JQ921438†	**KC610229**	x
	strouhali strouhali (Karaman, 1955)	UNTERACH_201009_1ID	Mondsee	AUSTRIA	47.79230	13.48452	Ha_153	Hb_130	st1	st1g	JQ921433†	**KC610230**	**KC610408**
	strouhali strouhali (Karaman, 1955)	UNTERACH_201009_3	Mondsee	AUSTRIA	47.79230	13.48452	Ha_154	Hb_130	st1	st1g	JQ921434†	**KC610231**	x
Proasellus synaselloides (Henry, 1963)	no	JABRON_201003_2	Sisteron	FRANCE	44.16769	5.92341	Ha_156	Hb_132	sy4	sy4g	**KC610487**	**KC610232**	x
	no	JABRON_201003_3	Sisteron	FRANCE	44.16769	5.92341	Ha_157	Hb_133	sy4	sy4g	**KC610488**	**KC610233**	x
	no	LAPENNE_201002_1ID	La penne sur l'ouveze	FRANCE	44.24240	5.22722	Ha_171	Hb_143	sy3	sy3g	JQ921449†	JQ921731†	JQ921998†
	no	LAPENNE_201002_3ID	La penne sur l'ouveze	FRANCE	44.24240	5.22722	Ha_172	Hb_144	sy3	sy3g	**KC610493**	**KC610244**	x
	no	LAPENNE_201002_4ID	La penne sur l'ouveze	FRANCE	44.24240	5.22722	Ha_173	Hb_145	sy3	sy3g	**KC610494**	**KC610245**	x
	no	MONTMAUR_201003_2	Montmaur	FRANCE	44.56021	5.89191	Ha_158	Hb_134	sy3	sy3g	**KC610489**	**KC610234**	x
	no	MONTMAUR_201003_3	Montmaur	FRANCE	44.56021	5.89191	Ha_159	Hb_135	sy4	sy4g	**KC610490**	**KC610235**	x
	no	SABRE_201101_1	Mazaugues	FRANCE	43.36607	5.92193	Ha_160	Hb_136	sy2	sy2g	**KC610491**	**KC610236**	x
	no	SENEZASS_201002_1ID	Senez	FRANCE	43.91530	6.40837	Ha_161	Hb_137	sy4	sy4g	JQ921440†	**KC610237**	**KC610409**
	no	SENEZASS_201002_2	Senez	FRANCE	43.91530	6.40837	Ha_162	Hb_138	sy4	sy4g	JQ921441†	**KC610238**	x
	no	THEOULE_200700_3	Meounes les montrieux	FRANCE	43.26361	5.95972	Ha_163	Hb_139	sy1	sy1g	**KC610492**	**KC610239**	x
	no	THEOULE_200700_4	Meounes les montrieux	FRANCE	43.26361	5.95972	Ha_164	Hb_139	sy1	sy1g	JQ921446†	**KC610240**	x

	no	THEOULE_200810_2ID	Meounes les montrieux	FRANCE	43.26361	5.95972	Ha_165	Hb_139	sy1	sy1g	JQ921445†	JQ921818†	JQ921996†
	no	VAUCLUS_200703_1	Fontaine de vaucluse	FRANCE	43.91861	5.13194	Ha_166	Hb_140	sy4	sy4g	JQ921442†	**KC610241**	x
	no	VAUCLUS_200703_3	Fontaine de vaucluse	FRANCE	43.91861	5.13194	Ha_167	Hb_141	sy4	sy4g	JQ921443†	JQ921817†	JQ921997†
	no	VAUCLUS_200703_4ID	Fontaine de vaucluse	FRANCE	43.91861	5.13194	Ha_168	Hb_140	sy4	sy4g	JQ921444†	**KC610242**	**KC610410**
	no	VIOLETTE_200707_1ID	Tenay	FRANCE	45.93667	5.53167	Ha_169	Hb_142	sy4	sy4g	JQ921447†	JQ921819†	JQ921999†
	no	VIOLETTE_200707_2ID	Tenay	FRANCE	45.93667	5.53167	Ha_170	Hb_142	sy4	sy4g	JQ921448†	**KC610243**	x
Proasellus walteri (Chappuis, 1948)	no	AZERGUES_200700_1	Ternand	FRANCE	45.92845	4.55026	Ha_177	Hb_149	wa3	wa3g	JQ921495†	JQ921825†	JQ922004†
	no	AZERGUES_200700_2ID	Ternand	FRANCE	45.92845	4.55026	Ha_178	Hb_149	wa3	wa3g	JQ921496†	**KC610246**	x
	no	BELMONT_200707_1ID	Belmont	FRANCE	47.00945	5.59472	Ha_179	Hb_150	wa2	wa2g	JQ921497†	**KC610247**	**KC610411**
	no	BURKH1_200905_1ex	Burkheim	GERMANY	48.09162	7.59935	Ha_180	Hb_151	wa1	wa1g	JQ921498†	KC610159§	x
	no	COLLIER_200804_1ID	Perreon	FRANCE	46.06694	4.57778	Ha_181	Hb_152	wa2	wa2g	JQ921510†	**KC610248**	**KC610412**
	no	COLLIER_200804_2	Perreon	FRANCE	46.06694	4.57778	Ha_182	Hb_153	wa2	wa2g	JQ921511†	**KC610249**	x
	no	COLLIER_200804_3	Perreon	FRANCE	46.06694	4.57778	Ha_183	Hb_152	wa2	wa2g	JQ921512†	**KC610250**	x
	no	COQUILLE_200807_1ID	Etalante	FRANCE	47.64500	4.76611	Ha_184	Hb_154	wa2	wa2g	JQ921499†	**KC610251**	**KC610413**
	no	COQUILLE_200807_2ID	Etalante	FRANCE	47.64500	4.76611	Ha_185	Hb_155	wa2	wa2g	JQ921500†	**KC610252**	x
	no	DROLABAT_201004_1	La batie des fonds	FRANCE	44.50862	5.62144	Ha_186*	Hb_156	wa3	wa3g	**KC610495**	**KC610253**	x
	no	DROLABAT_201004_2	La batie des fonds	FRANCE	44.50862	5.62144	Ha_186	Hb_156	wa3	wa3g	**KC610496**	x	x
	no	DROLIVRO_201004_1	Livron sur drome	FRANCE	44.76229	4.84090	Ha_188	Hb_157	wa3	wa3g	**KC610497**	**KC610254**	x
	no	DROLIVRO_201004_2	Livron sur drome	FRANCE	44.76229	4.84090	Ha_189	Hb_158	wa3	wa3g	**KC610498**	x	x
	no	LAPENNE_201002_2ex	La penne sur l'ouveze	FRANCE	44.24240	5.22722	Ha_190	Hb_159	wa3	wa3g	JQ921520†	JQ921828†	JQ922008†
	no	LONGAMON_200804_1ID	Longvic	FRANCE	47.28667	5.07944	Ha_191	Hb_160	wa2	wa2g	JQ921505†	**KC610255**	**KC610414**
	no	LONGAMON_200804_2	Longvic	FRANCE	47.28667	5.07944	Ha_192	Hb_161	wa2	wa2g	JQ921506†	**KC610256**	x
	no	LONGAMON_200804_3	Longvic	FRANCE	47.28667	5.07944	Ha_193	Hb_161	wa2	wa2g	JQ921507†	**KC610257**	x
	no	MONTBAR_200707_1	Montbarrey	FRANCE	47.01111	5.62972	Ha_194	Hb_162	wa2	wa2g	JQ921508†	JQ921827†	JQ922007†
	no	MONTBAR_200707_2	Montbarrey	FRANCE	47.01111	5.62972	Ha_195	Hb_163	wa2	wa2g	JQ921509†	**KC610258**	**KC610415**
	no	MONTJOU1_201002_1ID	Montjoux	FRANCE	44.50177	5.08935	Ha_196	Hb_164	wa3	wa3g	JQ921503†	**KC610259**	**KC610416**
	no	MONTJOU1_201002_2	Montjoux	FRANCE	44.50177	5.08935	Ha_197	Hb_165	wa3	wa3g	JQ921504†	**KC610260**	x
	no	ROSSFELD_200911_1ID	Rossfeld	FRANCE	48.33390	7.63116	Ha_198	Hb_166	wa1	wa1g	JQ921516†	JQ921829†	JQ922006†
	no	ROSSFELD_200911_2	Rossfeld	FRANCE	48.33390	7.63116	Ha_199	Hb_166	wa1	wa1g	JQ921517†	**KC610261**	**KC610417**
	no	ROUBSAUZ_201003_2	Sauzet	FRANCE	44.59126	4.81945	Ha_200	Hb_167	wa3	wa3g	**KC610499**	**KC610262**	x
	no	ROUBSAUZ_201003_3	Sauzet	FRANCE	44.59126	4.81945	Ha_201	Hb_167	wa3	wa3g	**KC610500**	**KC610263**	x
	no	SOUSPIRON_200700_2ID	L'isle sur la sorgue	FRANCE	43.90750	5.09000	Ha_202	Hb_168	wa3	wa3g	JQ921518†	**KC610264**	**KC610418**
	no	SOUSPIRON_200700_4	L'isle sur la sorgue	FRANCE	43.90750	5.09000	Ha_203	Hb_168	wa3	wa3g	JQ921519†	**KC610265**	x
	no	VERCLAUS_201002_1IDex	Verclause	FRANCE	44.38028	5.42590	Ha_204	Hb_169	wa3	wa3g	JQ921501†	JQ921826†	JQ922005†
	no	VERCLAUS_201002_2ex	Verclause	FRANCE	44.38028	5.42590	Ha_205	Hb_170	wa3	wa3g	JQ921502†	**KC610266**	x
	no	VIOLETTE_200707_1ID	Tenay	FRANCE	45.93667	5.53167	Ha_206	Hb_171	wa3	wa3g	JQ921513†	**KC610267**	x
	no	VIOLETTE_200707_2ID	Tenay	FRANCE	45.93667	5.53167	Ha_207	Hb_171	wa3	wa3g	JQ921514†	**KC610268**	x
	no	VIOLETTE_200707_3ID	Tenay	FRANCE	45.93667	5.53167	Ha_208	Hb_171	wa3	wa3g	JQ921515†	**KC610269**	x
Outgroup													
Proasellus albigensis (Magniez, 1965)		FOUSTER_200804_1	Cambon	FRANCE	43.91028	2.22278	Ha_1	Hb_1	T_40	G_14	JQ921028†	JQ921702†	JQ921886†
Proasellus aquaecalidae (Racovitza, 1922)		PEAUXCH_200811_1	Eaux bonnes	FRANCE	42.93861	-0.44028	Ha_2	Hb_2	T_46	G_28	JQ921035†	JQ921704†	JQ921888†
Proasellus aragonensis Henry & Magniez, 1992		BURGUI_200906_1ID	Burgui	SPAIN	42.70362	-1.01646	Ha_3	Hb_3	T_47	G_33	JQ921037†	JQ921705†	JQ921889†
Proasellus arnautovici (Remy, 1932)		VELIDAB_200909_1ID	Veli dab	MACEDONIA	41.00746	20.75565	Ha_4	Hb_4	T_48	G_34	JQ921043†	KC610091§	KC610338§
Proasellus arthrodilus (Braga, 1945)		LEGACAO_201010_2	Legacao	PORTUGAL	40.03177	-8.46979	Ha_5	Hb_5	T_37	G_35	JQ921045†	JQ921709†	JQ921892†
Proasellus assaforensis Afonso, 1988		ASSAFORA_201010_1ID	Assofora	PORTUGAL	38.90872	-9.42192	Ha_6	Hb_6	T_25	G_40	JQ921047†	JQ921710†	JQ921893†
Proasellus bardaunii Alouf, Henry & Magniez, 1982		QATTINE_201010_1ID	Tarchich	LEBANON	33.87570	35.79730	Ha_7	Hb_7	T_69	G_51	JQ921050†	JQ921711†	JQ921894†
Proasellus bellesi Henry & Magniez, 1982		TURON_200910_1ID	El burgo	SPAIN	36.78994	-4.94127	Ha_8	Hb_8	T_56	G_62	JQ921052†	JQ921712†	JQ921895†
Proasellus beroni Henry & Magniez, 1968		CASALUNA_200810_3	Lano	FRANCE	42.38750	9.25861	Ha_9	Hb_9	T_23	G_71	JQ921058†	JQ921714†	JQ921897†
Proasellus beticus Henry & Magniez, 1992		SUMIDORS_200910_1ID	Vallada	SPAIN	38.88048	-0.68871	Ha_10	Hb_10	T_62	G_15	JQ921061†	JQ921717†	JQ921899†
Proasellus boui Henry & Magniez, 1969		SAUVEGAR_000000_1ID	Sauve	FRANCE	43.94056	3.95000	Ha_11	Hb_11	T_76	G_18	JQ921064†	KC610092§	KC610339§
Proasellus cantabricus Henry & Magniez, 1968		RESCANO_200906_1ID	Mirones	SPAIN	43.29376	-3.69926	Ha_12	Hb_12	T_28	G_19	JQ921067†	JQ921719†	JQ921901†
Proasellus chappuisi Henry & Magniez, 1968		ARTZEGI_200906_2	Cigoitia	SPAIN	43.01902	-2.75427	Ha_56	Hb_46	T_36	G_36	JQ921121†	KC610094§	KC610341§
Proasellus chauvini Henry & Magniez, 1978		HAYTZA_200811_2	Saint etienne de baigorry	FRANCE	43.22222	-1.34528	Ha_57	Hb_47	T_41	G_37	JQ921124†	JQ921733†	JQ921915†
Proasellus claudei Henry & Magniez, 1996		VINCENT_201011_1ID	Saint etienne de baigorry	FRANCE	44.45945	1.29973	Ha_58	Hb_48	T_24	G_38	JQ921129†	JQ921735†	JQ921917†
Proasellus coiffaiti (Henry & Magniez, 1972)		NAVAILLO_200906_1ID	Navaillan	FRANCE	43.56129	-1.05223	Ha_59	Hb_49	T_68	G_39	JQ921131†	KC610095§	KC610342§
Proasellus comasi Henry & Magniez, 1982		VELEZ_200910_1	Vinuela	SPAIN	36.85189	-4.13523	Ha_60	Hb_50	T_57	G_41	JQ921134†	JQ921737†	JQ921919†
Proasellus coxalis (Dolfus, 1892)		RAAN_200800_3ID	Hellsingborg	SWEDEN	55.99900	12.76600	Ha_61	Hb_51	T_71	G_42	JQ921166†	KC610096§	KC610343§
Proasellus deminutus (Sket, 1959)		DOVJEZ_200910_4ex	Dovjez	SLOVENIA	46.11441	14.48526	Ha_62	Hb_52	T_4	G_43	JQ921176†	KC610097§	KC610344§

Proasellus dianae Pesce & Argano, 1985	CLITUNNO_200904_1ID	Camtello sul clitunno	ITALY	42.83293	12.76815	Ha_63	Hb_53	T_39	G_44	JQ921184†	JQ921746†	JQ921928†
Proasellus ebrensis Henry & Magniez, 1992	CERECEDA_200906_1	Cereceda	SPAIN	42.80023	-3.49401	Ha_64	Hb_54	T_31	G_45	JQ921187†	JQ921747†	JQ921929†
Proasellus escolai Henry & Magniez, 1982	DEIFONTE_200910_1ID	Deifontes	SPAIN	37.32921	-3.58740	Ha_65	Hb_55	T_55	G_46	JQ921192†	JQ921750†	JQ921932†
Proasellus espanoli Henry & Magniez, 1982	ALCAZAR_200910_1ID	Jerez del marquesado	SPAIN	37.17591	-3.16663	Ha_66	Hb_56	T_59	G_47	JQ921194†	JQ921751†	JQ921933†
Proasellus faesulanus Messana & Caselli, 1995	MUGNONE_200904_1ID	Pian di mugnona	ITALY	43.81933	11.29332	Ha_67	Hb_57	T_54	G_48	JQ921197†	KC610098§	KC610345§
Proasellus franciscoloi (Chappuis, 1955)	ORSO_200801_1ID	Ponte di nava	ITALY	44.11861	7.87444	Ha_68	Hb_58	T_11	G_49	JQ921202†	JQ921753†	JQ921935†
Proasellus sp. (type locality of P. gourbaultae)	CAUTAB_200910_1ex	Jarafuel	SPAIN	39.14380	-1.04686	Ha_69	Hb_59	T_52	G_50	JQ921111†	KC610099§	KC610346§
Proasellus grafi Henry & Magniez, 2003	SOURVALL_200906_1	Rasines	SPAIN	43.29843	-3.41996	Ha_70	Hb_60	T_30	G_52	JQ921207†	JQ921756†	JQ921938†
Proasellus granadensis Henry & Magniez, 2003	PILAS_200910_1IDex	Alhama de granada	SPAIN	37.04646	-4.06266	Ha_71	Hb_61	T_53	G_53	JQ921211†	KC610100§	KC610347§
Proasellus guipuzcoensis Henry & Magniez, 2003	UBAO_200906_2	Onate	SPAIN	43.00262	-2.40459	Ha_72	Hb_62	T_42	G_54	JQ921209†	KC610101§	KC610348§
Proasellus hercegovinensis (Karaman, 1933)	BJELUSNI_200008_1ID	Zavala	BOSNIA AND HERZEGOWINA	42.84516	17.97832	Ha_73	Hb_63	T_74	G_55	JQ921215†	JQ921760†	JQ921942†
Proasellus hermallensis (Arcangeli, 1938)	CHAMPP_200807_2ID	Yvoir	BELGIUM	50.31028	4.88866	Ha_74	Hb_64	T_43	G_56	JQ921217†	KC610102§	KC610349§
Proasellus ibericus (Braga, 1946)	BARRIO_201010_2	Barrio	PORTUGAL	41.84607	-8.56764	Ha_75	Hb_65	T_34	G_57	JQ921225†	KC610103§	KC610350§
Proasellus intermedius (Sket, 1965)	ZALOSCE_201008_3	Zalosce	SLOVENIA	45.88808	13.74331	Ha_76	Hb_66	T_6	G_58	JQ921234†	KC610104§	KC610351§
Proasellus istrianus (Stammer, 1932)	GLINSCIC_200701_1ID	Bagnoli	ITALY	45.62051	13.87228	Ha_77	Hb_67	T_50	G_59	JQ921239†	JQ921764†	JQ921946†
Proasellus jaloniacus Henry & Magniez, 1978	JALON_200910_1ID	Benichembla	SPAIN	38.75669	-0.10486	Ha_78	Hb_68	T_63	G_60	JQ921245†	JQ921765†	JQ921947†
Proasellus karamani (Remy, 1934)	ZALOMSKA_200305_1	Nevesinje	BOSNIA AND HERZEGOVINA	43.18116	18.12157	Ha_79	Hb_69	T_75	G_61	JQ921248†	JQ921766†	JQ921948†
Proasellus lagari Henry & Magniez, 1982	CRUZ_200910_1ID	Caravaca de la cruz	SPAIN	38.10056	-1.82796	Ha_80	Hb_70	T_58	G_63	JQ921249†	JQ921767†	JQ921949†
Proasellus lescherae Henry & Magniez, 1978	BERGANT_200910_1ID	Villores	SPAIN	40.66946	-0.19280	Ha_81	Hb_71	T_33	G_64	JQ921257†	JQ921768†	JQ921951†
Proasellus ligusticus Bodon & Argano, 1982	SORBA_200904_1ID	Monaglia	ITALY	44.24783	9.47847	Ha_82	Hb_72	T_38	G_65	JQ921275†	JQ921774†	JQ921956†
Proasellus margalefi Henry & Magniez, 1982	MEREMIL_200910_2	Gestalgar	SPAIN	39.59690	-0.84903	Ha_83	Hb_73	T_65	G_66	JQ921281†	KC610105§	KC610352§
Proasellus meijersae Henry & Magniez, 2003	MUNDO_200910_3	Riopar	SPAIN	38.48366	-2.36152	Ha_84	Hb_74	T_60	G_67	JQ921284†	JQ921777†	JQ921959†
Proasellus meridianus (Racovitza, 1919)	BARBOTT_200808_1	Thaire	FRANCE	46.05694	-0.97639	Ha_85	Hb_75	T_66	G_68	JQ921285†	KC610106§	KC610353§
Proasellus micropectinatus Baratti & Messana, 1990	MARGHE_200904_2	Sesta godano	ITALY	44.27222	9.65710	Ha_86	Hb_76	T_51	G_69	JQ921326†	KC610107§	KC610354§
Proasellus navarrensis Henry & Magniez, 2003	NACEDERO_200906_1ID	Larraun	SPAIN	42.97672	-1.91840	Ha_87	Hb_77	T_35	G_70	JQ921328†	JQ921783†	JQ921964†
Proasellus ortizi Henry & Magniez, 1992	JIVERO2_200906_3	Ozana	SPAIN	43.30385	-3.57186	Ha_93	Hb_81	T_29	G_72	JQ921364†	JQ921798†	JQ921978†
Proasellus oviedensis Henry & Magniez, 2003	QUINTANA_200906_1	La pereda	SPAIN	43.40031	-4.77120	Ha_94	Hb_82	T_64	G_73	JQ921365†	JQ921799†	JQ921979†
Proasellus parvulus (Sket, 1960)	OTOVEC_201009_1ID	Crnomelj	SLOVENIA	45.59231	15.16690	Ha_95	Hb_83	T_73	G_74	JQ921366†	KC610108§	KC610355§
Proasellus pavani (Arcangeli, 1942)	STERPO_201008_3	Bertiolo	ITALY	45.90291	13.04147	Ha_96	Hb_84	T_70	G_75	JQ921371†	KC610109§	KC610356§
Proasellus racovitzai Henry & Magniez, 1972	BERNATAS_200811_3ID	Arbas	FRANCE	42.97306	0.89889	Ha_97	Hb_85	T_61	G_76	JQ921372†	JQ921802†	JQ921982†
Proasellus rectangulatus Afonso, 1982	MONTEMOR_201010_2	Montemor o novo	PORTUGAL	38.65417	-8.22641	Ha_98	Hb_86	T_26	G_77	JQ921374†	KC610110§	KC610357§
Proasellus rectus Afonso, 1982	EVORA_201010_1ID	Evora	PORTUGAL	38.60359	-7.87318	Ha_99	Hb_87	T_27	G_78	JQ921376†	JQ921804†	JQ921984†
Proasellus remyi (Monod, 1932)	OSOJ2_200909_1ID	Pod osoj	MACEDONIA	40.95107	20.77643	Ha_100	Hb_88	T_49	G_16	JQ921379†	KC610111§	KC610358§
Proasellus rouchi Henry, 1980	MESCLA_200904_5	Villars sur var	FRANCE	43.91250	7.18583	Ha_101	Hb_89	T_18	G_17	JQ921385†	KC610112§	KC610359§
Proasellus slovenicus (Sket, 1957)	JAMAPOD_200409_1	Precna	SLOVENIA	45.81886	15.09865	Ha_129	Hb_111	T_67	G_20	JQ921413†	JQ921810†	JQ921990†
Proasellus spelaeus (Racovitza, 1922)	ETXANKO_200811_1	Aussurucq	FRANCE	43.12389	-0.97139	Ha_130	Hb_112	T_44	G_21	KC610466§	KC610113§	KC610360§
Proasellus stocki Henry & Magniez, 2003	SAJA_201010_1ID	Cabezon de la sal	SPAIN	43.12104	-4.29230	Ha_131	Hb_113	T_32	G_22	JQ921116†	JQ921730†	JQ921913†
Proasellus valdensis (Chappuis, 1948)	MOUDOUBS_201006_1	Mouthe	FRANCE	46.70497	6.20930	Ha_174	Hb_146	T_10	G_25	KC610478§	KC610149§	x
Proasellus vandeli Magniez & Henry, 1969	UTHURRIA_201005_2	Viodos-abense-de-bas	FRANCE	43.23386	-0.91735	Ha_175	Hb_147	T_45	G_26	JQ921490†	KC610158§	KC610368§
Proasellus vulgaris (Sket, 1965)	DOVJEZ_200910_2ex	Dovjez	SLOVENIA	46.11441	14.48526	Ha_176	Hb_148	T_5	G_27	JQ921493†	JQ921824†	JQ922003†
Caecidotea kenki (Bowman, 1967)	PIMMIT1_200902_2ID	Fairfax	USA	38.92917	-77.11806	Ha_209	Hb_172	T_3	G_29	JQ921575†	JQ921691†	JQ921877†
Gallasellus heilyi (Legrand, 1956)	PUITD_200912_1ID	Le thou	FRANCE	46.08472	-0.92528	Ha_210	Hb_173	T_72	G_30	JQ921582†	JQ921694†	JQ921880†
Stenasellus breuili Racovitza, 1924	ALLI_200906_1ID	Larraun	SPAIN	42.98775	-1.89284	Ha_211	Hb_174	T_1	G_31	JQ921607†	JQ921830†	x
Stenasellus racovitzai Razzauti, 1925	CASAMOZA_200810_1	Lucciana	FRANCE	42.51861	9.44222	Ha_212	Hb_175	T_2	G_32	JQ921621†	JQ921836†	JQ922011†

* Denote individuals used to represent a given haplotype within phylogenetic trees used for testing monophyly.

† Morvan C, Malard F, Paradis E, Lefébure T, Konecny-Dupré L, Douady CJ (2013) Timetree of Aselloidea reveals species diversification dynamics in groundwater. *Systematic Biology*, doi: 10.1093/sysbio/syt015

§ Eme D, Malard F, Colson-Proch C, Jean P, Calvignac S, Konecny-Dupré L, Hervant F, Douady CJ (2013) Integrating phylogeography, physiology, and habitat modelling to explore species range determinants. Submitted

Table S2 Molecular data acquisition. Primer pairs. F: forward, R : reverse.

Gene	Primer name	F/R	5'-3' Sequence	Reference
COI	COI-LCO1490	F	GGTCAACAAATCATAAAGATATTGG	Folmer *et al.* 1994*
COI	COI-HCO2198	R	TAAACTTCAGGGTGACCAAAAAATCA	Folmer *et al.* 1994
COI	COI-Pcava-F1	F	CTTAGTATACTGATCCGCACAGAGCTTGG	Morvan et al. 2013†
COI	COI-Pcava-R1	R	GGGGTCAAAAAATGAAGTATTTAAGTTAC	Morvan *et al.* 2013
COI	COI-Pstrou-R1	R	AGGTGTTGGTACAGAATTGGGTCCCCTC	Calvignac *et al.* 2011§
COI	COI-Pwalt-F1	F	CACCCTATATTTTATCTTCGGCGCCTG	Calvignac *et al.* 2011
COI	COI-Psyna-F1	F	GCTTGAGCAGGCAGAGTCGGAACTGC	Calvignac *et al.* 2011
COI	COI-Psyna-R1	R	GTATTTAAGTTACGGTCAGTTAATAGTATGGTGATAG	Calvignac *et al.* 2011
COI	F1-FS-COI	F	ACCAGGCAGTTTTATTGGCAA	Morvan *et al.* 2013
COI	R2-FS-COI	R	GTATTTAGGTTGCGATCTGTAAGTA	Morvan *et al.* 2013
COI	COI_StenaR719	R	TTTATTAGACTCTTGACTAACGATATGAG	Morvan *et al.* 2013
COI	H6951	R	ACACCTGTTAACCCTCCTAC	Theisen et al. 1995¥
COI	COI_StenaR508	R	AAAGCTGTTAAGAATGCAGACCAGAC	Morvan *et al.* 2013
COI	COI_stena_F1	F	TTAAGWATATTAATTCGTTCWGARYTAGG	Morvan *et al.* 2013
16S	16Sar-Dr	F	CGCCTGTTTAACAAAAACAT	Palumbi & Benzie, 1991ł
16S	16Sbr	R	CCGGTCTGAACTCAAGATCACGT	Palumbi, 1996‡
16S	16S-Proa-F1	F	CCTATGAGTCGTTTAAATGGCCGCA	Calvignac *et al.* 2011
16S	16S_Stena_R1	R	CGTGGAAGTTTAATAGTCGAACAGAC	Calvignac *et al.* 2011
16S	16S_Stena_F1	F	CCTATGAATTATTTTAATGGCCGCAG	Calvignac *et al.* 2011
16S	16S-Pwalt-F1	F	CCCTATGAATTATTTTAATGGCCGCA	Morvan *et al.* 2013
16S	16S-Psy.wa.st-R1	R	CCCAATTCAACATCGAGGTCGTAAG	Morvan *et al.* 2013
28S	28S_stena_R1	R	CCTATACCCAGCTCTGACGATCG	Morvan *et al.* 2013
28S	28S_des1	R	AGGGAAACTTCGGAGGGAACC	Morvan *et al.* 2013
28S	28S_lev2	F	CAAGTACCGTGAGGGAAAGTT	Morvan *et al.* 2013
28S	28S_stena_F1_A	F	GAGTTCAAGAGAACGTGAAACTACCA	Morvan *et al.* 2013
28S	28S_univ_R2	R	GTCCGTAGCGATTCTGACGTGC	Morvan *et al.* 2013

*Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*,**3**, 294-299.
†Morvan C, Malard F, Paradis E, Lefébure T, Konecny-Dupré L, Douady CJ (2013) Timetree of Aselloidea reveals species diversification dynamics in groundwater. *Systematic Biology*, doi: 10.1093/sysbio/syt015
§Calvignac S, Konecny L, Malard F, Douady CJ (2011) Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). *Mitochondrion*, **11**, 246-25.
¥Theisen BF, Christensen B, Arctander P (1995) Origin of clonal diversity in triploid parthenogenetic *Trichoniscus pusillus pusillus* (Isopoda, Crustacea) based upon allozyme and nucleotide sequence data. *Journal of Evolutionary Biology*, **8**, 71-80.
łPalumbi SR, Benzie J (1991) Large mitochondrial DNA differences among morphologically similar Penaeid shrimp. *Molecular Marine Biology and Biotechnology*, **1**, 27-34.
‡Palumbi SR (1996) In Molecular Systematics, (eds Hillis DM, Moritz C, Mable BK), pp 205-248, Sinauer Associates, Sunderland, MA

DNA was extracted from specimens following a classic chloroform DNA extraction protocol used by Calvignac *et al.* (2011). Then, we amplified DNA with primers targeting the mitochondrial cytochrome oxidase subunit I (COI) gene, 16S mitochondrial rDNA gene and 28S nuclear rDNA gene (Calvignac *et al.* 2011; Morvan *et al.* 2013). All primer pairs were described in table 1. PCR reactions were managed following an optimized protocol (Morvan et al. 2013) using a Taq polymerase concentration of 0.04 U. The settings of PCR reaction were as follow: (i) one step of 2 min at 94°C, (ii) 40 (COI) or 35 (16S and 28S) cycles of 30s at 94°C, 30 sec at 48°C (COI), 53°C (16S) or 62°C (28S), 30 sec at 72°C and (iii) one step of 10 min at 72°C. To avoid misleading inclusion of nuclear mitochondrial pseudogenes (*numts*) in COI and 16S datasets, we followed Calvignac *et al.*(2010) and combined for each locality three methods: different primers pairs (table 1), long–range amplification and pre–PCR dilution of genomic DNA. Concerning long-range amplification, we used 0.09 U Taq

DNA polymerase concentration instead of 0.15 U used by Calvignac *et al.* (2010). Additionally, to detect putative paralogs when characterizing 28S fragments, we also used different primer pairs (table 1). Sanger sequencing was performed by service providers (GATC Biotech; Konstanz, Germany; Eurofins MWG Operon, Ebersberg, Germany; BIOFIDAL; Vaulx–en–Velin, France). Chromatograms were visualized using FinchTV (Geospiza, Inc.; Seattle, WA, USA; http://www.geospiza.com).

Annexe 6 : Matériel supplémentaire Article 5

Journal of Biogeography

SUPPORTING INFORMATION

Integrating phylogeography, physiology, and habitat modelling to explore species range determinants

David Eme, Florian Malard, Céline Colson-Proch, Pauline Jean, Sébastien Calvignac, Lara Konecny-Dupré, Frédéric Hervant and Christophe J. Douady

Appendix S1 Sampling sites, morphospecies, accession numbers and priors used in BEAST for assessing monophyly of *Proasellus valdensis* and testing for the presence of cryptic species.

Table S1 Sampling sites, morphospecies, accession numbers and priors used in BEAST for testing monophyly of *Proasellus valdensis* and presence of cryptic species. Locality, nearest township; Mount., associated mountains; Co., iso country codes; Lat., latitude in decimal degrees; Long., longitude in decimal degrees; HA, haplotype number used for testing the monophyly; HB, intraspecific haplotypes of *COI* gene; HC, intraspecific haplotypes of *16S* gene; COI_SP, evolutionary unit number as delimited with threshold and GMYC methods; ANCOI, accession number for *COI* sequence; AN16S, accession number for *16S* sequence; AN28S, accession number for *28S* sequence. Sequences marked in bold are from Morvan *et al.* (2013).

Species	Individual	Locality	Mount.	Co.	Lat.	Long.	HA	HB	HC	COI_SP	ANCOI	AN16S	AN28S
Proasellus valdensis (Chappuis, 1948)	BALME_200709_1	La Balme-les-Grottes	Crémieu	FR	45.852	5.339	Ha_62†	Hb_4	Hc_4	unit24	**JQ921452**	KC610115	KC610362
	BALME_200709_19ml	La Balme-les-Grottes	Crémieu	FR	45.852	5.339	Ha_63	Hb_4	Hc_5	unit24	**JQ921450**	KC610116	×
	BALME_200709_20ml	La Balme-les-Grottes	Crémieu	FR	45.852	5.339	Ha_62	Hb_4	×	unit24	**JQ921451**	×	×
	BENOITE_000000_1	Arith	Bauges	FR	45.708	6.042	Ha_65	Hb_9	Hc_7	unit24	KC610467	KC610117	×
	BENOITE_000000_2	Arith	Bauges	FR	45.708	6.042	Ha_66	Hb_8	Hc_7	unit24	KC610468	KC610118	×
	BENOITE_000000_3	Arith	Bauges	FR	45.708	6.042	Ha_67	Hb_9	Hc_7	unit24	**JQ921453**	KC610119	×
	BERNARD_200802_1	Samoëns	Haut-Giffre	FR	46.102	6.780	Ha_68†	Hb_11	Hc_12	unit24	**JQ921468**	KC610120	×
	BERNARD_200802_2	Samoëns	Haut-Giffre	FR	46.102	6.780	Ha_68	Hb_11	Hc_12	unit24	KC610469	KC610121	×
	BERNARD_200802_3	Samoëns	Haut-Giffre	FR	46.102	6.780	Ha_70	Hb_11	Hc_12	unit24	KC610470	KC610122	×
	CAVALE_200511_27ml	Les Déserts	Bauges	FR	45.665	5.990	Ha_71	Hb_9	Hc_7	unit24	**JQ921454**	KC610123	×
	CAVALE_200511_28ml	Les Déserts	Bauges	FR	45.665	5.990	Ha_72	Hb_9	Hc_15	unit24	**JQ921455**	KC610124	×
	CAVALE_200511_29ml	Les Déserts	Bauges	FR	45.665	5.990	Ha_73	Hb_9	Hc_15	unit24	**JQ921456**	KC610125	×
	CHRISTOP_200511_1ID	Saint-Christophe	Chartreuse	FR	45.438	5.438	Ha_74	Hb_16	Hc_16	unit24	KC610471	KC610126	×
	CLAUDE_200504_24ml	Saint-Claude	Jura	FR	46.364	5.878	Ha_75	Hb_17	Hc_17	unit24	**JQ921484**	KC610127	×

Species	Individual	Locality	Mount.	Co.	Lat.	Long.	HA	HB	HC	COI_SP	ANCOI	AN16S	AN28S
	CLAUDE_200504_25ml	Saint-Claude	Jura	FR	46.364	5.878	Ha_76	Hb_17	Hc_18	unit24	KC610472	KC610128	×
	CLAUDE_200504_26ml	Saint-Claude	Jura	FR	46.364	5.878	Ha_77	Hb_17	Hc_19	unit24	**JQ921485**	KC610129	×
	CUVES_200906_1ID	Sassenage	Vercors	FR	45.209	5.659	Ha_78	Hb_20	Hc_22	unit24	**JQ921480**	KC610130	KC610363
	CUVES_200906_2	Sassenage	Vercors	FR	45.209	5.659	Ha_79	Hb_20	Hc_22	unit24	**JQ921481**	KC610131	×
	CUVES_200906_3	Sassenage	Vercors	FR	45.209	5.659	Ha_80	Hb_20	Hc_22	unit24	**JQ921482**	KC610132	×
	DOUVERA_200908_1ID	La Pesse	Jura	FR	46.318	5.834	Ha_81	Hb_25	Hc_24	unit24	**JQ921469**	KC610133	×
	DOUVERA_200908_2	La Pesse	Jura	FR	46.318	5.834	Ha_82	Hb_24	Hc_24	unit24	**JQ921470**	KC610134	×
	DOUVERA_200908_3	La Pesse	Jura	FR	46.318	5.834	Ha_83	Hb_25	Hc_24	unit24	**JQ921471**	KC610135	×
	FALCON_200806_1	La Burbanche	Jura	FR	45.842	5.547	Ha_84	Hb_26	Hc_26	unit24	**JQ921458**	KC610136	×
	FALCON_200806_2	La Burbanche	Jura	FR	45.842	5.547	Ha_85	Hb_28	Hc_28‡	unit24	KC610473	KC610137	×
	FALCON_200806_3	La Burbanche	Jura	FR	45.842	5.547	Ha_86	Hb_28	Hc_28‡	unit24	KC610474	KC610138	×
	FARDELE_200704_1f	Sixt-Fer-à-Cheval	Haut-Giffre	FR	46.025	6.773	Ha_87	Hb_29	Hc_29	unit24	**JQ921462**	KC610139	×
	FARDELE_200704_2	Sixt-Fer-à-Cheval	Haut-Giffre	FR	46.025	6.773	Ha_88	Hb_29	Hc_29	unit24	KC610475	KC610140	×
	FARDELW_200602_17ml	Sixt-Fer-à-Cheval	Haut-Giffre	FR	46.025	6.773	Ha_89	Hb_31	Hc_31	unit24	**JQ921459**	KC610141	×
	FARDELW_200602_18ml	Sixt-Fer-à-Cheval	Haut-Giffre	FR	46.025	6.773	Ha_90	Hb_29	Hc_32	unit24	**JQ921460**	KC610142	×
	FARDELW_200602_22ml	Sixt-Fer-à-Cheval	Haut-Giffre	FR	46.025	6.773	Ha_91	Hb_9	Hc_31	unit24	**JQ921461**	KC610143	×
	FONTAGNY_200602_2	Sixt-Fer-à-Cheval	Haut-Giffre	FR	46.071	6.805	Ha_92	Hb_34	×	unit24	KC610476	×	×
	FOULESC_200710_1ID	Saint-Claude	Jura	FR	46.377	5.896	Ha_93	Hb_35	Hc_35	unit24	**JQ921463**	KC610144	×
	FOULESC_200710_3ID	Saint-Claude	Jura	FR	46.377	5.896	Ha_94	Hb_36	Hc_36	unit24	**JQ921464**	KC610145	×

Species	Individual	Locality	Mount.	Co.	Lat.	Long.	HA	HB	HC	COI_SP	ANCOI	AN16S	AN28S
	FURON_200808_1	Engins	Vercors	FR	45.167	5.612	Ha_95	Hb_37	Hc_37	unit24	**JQ921457**	KC610146	KC610364
	HUGBIS_200709_2	Injoux	Jura	FR	46.032	5.769	Ha_96†	Hb_38	Hc_38	unit24	**JQ921465**	**JQ921820**	**JQ922000**
	HUGBIS_200709_3	Injoux	Jura	FR	46.032	5.769	Ha_96	Hb_38	Hc_38	unit24	**JQ921466**	KC610147	×
	HUGBIS_200709_4	Injoux	Jura	FR	46.032	5.769	Ha_96	Hb_38	×	unit24	**JQ921467**	×	×
	MOREARCE_201006_1	Morez	Jura	FR	46.512	6.028	Ha_99	Hb_36	Hc_36	unit24	KC610477	KC610148	×
	MOUDOUBS_201006_1	Mouthe	Jura	FR	46.705	6.209	Ha_100	Hb_35	Hc_35	unit24	KC610478	KC610149	×
	PLAGNE_200511_33ml	Entremont-le-Vieux	Chartreuse	FR	45.440	5.904	Ha_101	Hb_9	Hc_16	unit24	**JQ921472**	KC610150	KC610365
	PLAGNE_200511_34ml	Entremont-le-Vieux	Chartreuse	FR	45.440	5.904	Ha_102	Hb_9	Hc_44‡	unit24	**JQ921473**	KC610151	×
	PLEUREUS_200509_1	Sixt-Fer-à-Cheval	Haut-Giffre	FR	46.012	6.759	Ha_103	Hb_51	Hc_46	unit24	**JQ921476**	KC610152	×
	PLEUREUS_200509_13ml	Sixt-Fer-à-Cheval	Haut-Giffre	FR	46.012	6.759	Ha_104†	Hb_51	Hc_46	unit24	**JQ921474**	KC610153	KC610366
	PLEUREUS_200509_14ml	Sixt-Fer-à-Cheval	Haut-Giffre	FR	46.012	6.759	Ha_104	Hb_51	Hc_46	unit24	**JQ921475**	KC610154	×
	SALLES_200509_15ml	Sixt-Fer-à-Cheval	Haut-Giffre	FR	46.001	6.754	Ha_106	Hb_48	Hc_48‡	unit24	**JQ921477**	KC610155	×
	SALLES_200509_16ml	Sixt-Fer-à-Cheval	Haut-Giffre	FR	46.001	6.754	Ha_107	Hb_49	Hc_48‡	unit24	**JQ921478**	KC610156	×
	SALLES_200509_7	Sixt-Fer-à-Cheval	Haut-Giffre	FR	46.001	6.754	Ha_108	Hb_50	×	unit24	**JQ921479**	×	×
	SCESIXT_200704_1	Sixt-Fer-à-Cheval	Haut-Giffre	FR	46.057	6.784	Ha_109	Hb_51	Hc_46	unit24	**JQ921483**	KC610157	KC610367
Caecidotea kenki (Bowman, 1967)	PIMMIT1_200902_2ID	Fairfax		US	38.929	−77.118	Ha_113			unit9	**JQ921575**	**JQ921691**	**JQ921877**
Gallasellus heilyi (Legrand, 1956)	PUITD_200912_1ID	Le Thou		FR	46.085	−0.925	Ha_114			unit4	**JQ921582**	**JQ921694**	**JQ921880**
P. albigensis (Magniez, 1965)	FOUSTER_200804_1	Cambon		FR	43.910	2.223	Ha_1			unit37	**JQ921028**	**JQ921702**	**JQ921886**

Species	Individual	Locality	Mount.	Co.	Lat.	Long.	HA	HB	HC	COI_SP	ANCOI	AN16S	AN28S
P. aquaecalidae (Racovitza, 1922)	PEAUXCH_200811_1	Eaux-Bonnes		FR	42.939	−0.440	Ha_2			unit43	**JQ921035**	**JQ921704**	**JQ921888**
P. aragonensis Henry & Magniez, 1992	BURGUI_200906_1ID	Burgui		ES	42.704	−1.016	Ha_3			unit44	**JQ921037**	**JQ921705**	**JQ921889**
P. arnautovici (Remy, 1932)	VELIDAB_200909_1ID	Veli Dab		MK	41.007	20.756	Ha_4			unit48	**JQ921043**	KC610091	KC610338
P. arthrodilus (Braga, 1945)	LEGACAO_201010_2	Legacao		PT	40.032	−8.470	Ha_5			unit28	**JQ921045**	**JQ921709**	**JQ921892**
P. assaforensis Afonso, 1988	ASSAFORA_201010_1ID	Assofora		PT	38.909	−9.422	Ha_6			unit36	**JQ921047**	**JQ921710**	**JQ921893**
P. bardaunii Alouf, Henry & Magniez, 1982	QATTINE_201010_1ID	Tarchich		LB	33.876	35.797	Ha_7			unit16	**JQ921050**	**JQ921711**	**JQ921894**
P. bellesi Henry & Magniez, 1982	TURON_200910_1ID	El Burgo		ES	36.790	−4.941	Ha_8			unit63	**JQ921052**	**JQ921712**	**JQ921895**
P. beroni Henry & Magniez, 1968	CASALUNA_200810_3	Lano		FR	42.388	9.259	Ha_9			unit50	**JQ921058**	**JQ921714**	**JQ921897**
P. beticus Henry & Magniez, 1992	SUMIDORS_200910_1ID	Vallada		ES	38.880	−0.689	Ha_10			unit54	**JQ921061**	**JQ921717**	**JQ921899**
P. boui Henry & Magniez, 1969	SAUVEGAR_000000_1ID	Sauve		FR	43.941	3.950	Ha_11			unit3	**JQ921064**	KC610092	KC610339
P. cantabricus Henry & Magniez, 1968	RESCANO_200906_1ID	Mirones		ES	43.294	−3.699	Ha_12			unit34	**JQ921067**	**JQ921719**	**JQ921901**
P. cavaticus (Leydig, 1871)	BURKH3_200905_1ID	Burkheim		DE	48.095	7.593	Ha_13			unit18	KC610465	KC610093	KC610340
P. chappuisi Henry & Magniez, 1968	ARTZEGI_200906_2	Cigoitia		ES	43.019	−2.754	Ha_14			unit27	**JQ921121**	KC610094	KC610341
P. chauvini Henry & Magniez, 1978	HAYTZA_200811_2	Saint-Etienne-de-Baïgorry		FR	43.222	−1.345	Ha_15			unit45	**JQ921124**	**JQ921733**	**JQ921915**
P. claudei Henry & Magniez, 1996	VINCENT_201011_1ID	Saint-Vincent-Rive-d'Olt		FR	44.459	1.300	Ha_16			unit67	**JQ921129**	**JQ921735**	**JQ921917**
P. coiffaiti (Henry & Magniez, 1972)	NAVAILLO_200906_1ID	Navaillan		FR	43.561	−1.052	Ha_17			unit13	**JQ921131**	KC610095	KC610342

Species	Individual	Locality	Mount.	Co.	Lat.	Long.	HA	HB	HC	COI_SP	ANCOI	AN16S	AN28S
P. comasi Henry & Magniez, 1982	VELEZ_200910_1	Vinuela		ES	36.852	−4.135	Ha_18			unit59	**JQ921134**	**JQ921737**	**JQ921919**
P. coxalis (Dolfus, 1892)	RAAN_200800_3ID	Hellsingborg		SE	55.999	12.766	Ha_19			unit8	**JQ921166**	KC610096	KC610343
P. deminutus (Sket, 1959)	DOVJEZ_200910_4ex	Dovjez		SI	46.114	14.485	Ha_20			unit11	**JQ921176**	KC610097	KC610344
P. dianae Pesce & Argano, 1985	CLITUNNO_200904_1ID	Camtello sul Clitunno		IT	42.833	12.768	Ha_21			unit52	**JQ921184**	**JQ921746**	**JQ921928**
P. ebrensis Henry & Magniez, 1992	CERECEDA_200906_1	Cereceda		ES	42.800	−3.494	Ha_22			unit32	**JQ921187**	**JQ921747**	**JQ921929**
P. escolai Henry & Magniez, 1982	DEIFONTE_200910_1ID	Deifontes		ES	37.329	−3.587	Ha_23			unit60	**JQ921192**	**JQ921750**	**JQ921932**
P. espanoli Henry & Magniez, 1982	ALCAZAR_200910_1ID	Jerez del Marquesado		ES	37.176	−3.167	Ha_24			unit61	**JQ921194**	**JQ921751**	**JQ921933**
P. faesulanus Messana & Caselli, 1995	MUGNONE_200904_1ID	Pian di Mugnona		IT	43.819	11.293	Ha_25			unit58	**JQ921197**	KC610098	KC610345
P. franciscoloi (Chappuis, 1955)	ORSO_200801_1ID	Ponte di Nava		IT	44.119	7.874	Ha_26			unit23	**JQ921202**	**JQ921753**	**JQ921935**
P. grafi Henry & Magniez, 2003	SOURVALL_200906_1	Rasines		ES	43.298	−3.420	Ha_28			unit30	**JQ921207**	**JQ921756**	**JQ921938**
P. granadensis Henry & Magniez, 2003	PILAS_200910_1IDex	Alhama de Granada		ES	37.046	−4.063	Ha_29			unit64	**JQ921211**	KC610100	KC610347
P. guipuzcoensis Henry & Magniez, 2003	UBAO_200906_2	Onate		ES	43.003	−2.405	Ha_30			unit46	**JQ921209**	KC610101	KC610348
P. hercegovinensis (Karaman, 1933)	BJELUSNI_200008_1ID	Zavala		BA	42.845	17.978	Ha_31			unit2	**JQ921215**	**JQ921760**	**JQ921942**
P. hermallensis (Arcangeli, 1938)	CHAMPP_200807_2ID	Yvoir		BE	50.310	4.889	Ha_32			unit40	**JQ921217**	KC610102	KC610349
P. ibericus (Braga, 1946)	BARRIO_201010_2	Barrio		PT	41.846	−8.568	Ha_33			unit47	**JQ921225**	KC610103	KC610350
P. intermedius (Sket, 1965)	ZALOSCE_201008_3	Zalosce		SI	45.888	13.743	Ha_34			unit10	**JQ921234**	KC610104	KC610351
P. istrianus (Stammer, 1932)	GLINSCIC_200701_1ID	Bagnoli		IT	45.621	13.872	Ha_35			unit25	**JQ921239**	**JQ921764**	**JQ921946**

Species	Individual	Locality	Mount.	Co.	Lat.	Long.	HA	HB	HC	COI_SP	ANCOI	AN16S	AN28S
P. jaloniacus Henry & Magniez, 1978	JALON_200910_1ID	Benichembla		ES	38.757	−0.105	Ha_36			unit55	**JQ921245**	**JQ921765**	**JQ921947**
P. karamani (Remy, 1934)	ZALOMSKA_200305_1	Nevesinje		BA	43.181	18.122	Ha_37			unit15	**JQ921248**	**JQ921766**	**JQ921948**
P. lagari Henry & Magniez, 1982	CRUZ_200910_1ID	Caravaca de la Cruz		ES	38.101	−1.828	Ha_38			unit65	**JQ921249**	**JQ921767**	**JQ921949**
P. lescherae Henry & Magniez, 1978	BERGANT_200910_1ID	Villores		ES	40.669	−0.193	Ha_39			unit56	**JQ921257**	**JQ921768**	**JQ921951**
P. ligusticus Bodon & Argano, 1982	SORBA_200904_1ID	Monaglia		IT	44.248	9.478	Ha_40			unit51	**JQ921275**	**JQ921774**	**JQ921956**
P. margalefi Henry & Magniez, 1982	MEREMIL_200910_2	Gestalgar		ES	39.597	−0.849	Ha_41			unit53	**JQ921281**	KC610105	KC610352
P. meijersae Henry & Magniez, 2003	MUNDO_200910_3	Riopar		ES	38.484	−2.362	Ha_42			unit66	**JQ921284**	**JQ921777**	**JQ921959**
P. meridianus (Racovitza, 1919)	BARBOTT_200808_1	Thaire		FR	46.057	−0.976	Ha_43			unit29	**JQ921285**	KC610106	KC610353
P. micropectinatus Baratti & Messana, 1990	MARGHE_200904_2	Sesta Godano		IT	44.272	9.657	Ha_44			unit26	**JQ921326**	KC610107	KC610354
P. navarrensis Henry & Magniez, 2003	NACEDERO_200906_1ID	Larraun		ES	42.977	−1.918	Ha_45			unit35	**JQ921328**	**JQ921783**	**JQ921964**
P. nolli (Karaman, 1952)	ROSSFELD_200911_1ID	Rossfeld		FR	48.334	7.631	Ha_46			unit1	**JQ921332**	**JQ921784**	**JQ921965**
P. ortizi Henry & Magniez, 1992	JIVERO2_200906_3	Ozana		ES	43.304	−3.572	Ha_47			unit33	**JQ921364**	**JQ921798**	**JQ921978**
P. oviedensis Henry & Magniez, 2003	QUINTANA_200906_1	La Pereda		ES	43.400	−4.771	Ha_48			unit57	**JQ921365**	**JQ921799**	**JQ921979**
P. parvulus (Sket, 1960)	OTOVEC_201009_1ID	Crnomelj		SI	45.592	15.167	Ha_49			unit14	**JQ921366**	KC610108	KC610355
P. pavani (Arcangeli, 1942)	STERPO_201008_3	Bertiolo		IT	45.903	13.041	Ha_50			unit5	**JQ921371**	KC610109	KC610356
P. racovitzai Henry & Magniez, 1972	BERNATAS_200811_3ID	Arbas		FR	42.973	0.899	Ha_51			unit68	**JQ921372**	**JQ921802**	**JQ921982**

Species	Individual	Locality	Mount.	Co.	Lat.	Long.	HA	HB	HC	COI_SP	ANCOI	AN16S	AN28S
P. rectangulatus Afonso, 1982	MONTEMOR_201010_2	Montemor o Novo		PT	38.654	−8.226	Ha_52			unit38	**JQ921374**	KC610110	KC610357
P. rectus Afonso, 1982	EVORA_201010_1ID	Evora		PT	38.604	−7.873	Ha_53			unit39	**JQ921376**	**JQ921804**	**JQ921984**
P. remyi (Monod, 1932)	OSOJ2_200909_1ID	Pod Osoj		MK	40.951	20.776	Ha_54			unit49	**JQ921379**	KC610111	KC610358
P. rouchi Henry, 1980	MESCLA_200904_5	Villars-sur-Var		FR	43.913	7.186	Ha_55			unit19	**JQ921385**	KC610112	KC610359
P. slavus (Remy, 1948)	GLANBACH_201009_1ID	Salzburg		AT	47.810	13.022	Ha_56			unit1	**JQ921400**	**JQ921807**	**JQ921987**
P. slovenicus (Sket, 1957)	JAMAPOD_200409_1	Precna		SI	45.819	15.099	Ha_57			unit17	**JQ921413**	**JQ921810**	**JQ921990**
P. sp. (type locality of *P. gourbaultae*)	CAUTAB_200910_1ex	Jarafuel		ES	39.144	−1.047	Ha_27			unit62	**JQ921111**	KC610099	KC610346
P. spelaeus (Racovitza, 1922)	ETXANKO_200811_1	Aussurucq		FR	43.124	−0.971	Ha_58			unit41	KC610466	KC610113	KC610360
P. stocki Henry & Magniez, 2003	SAJA_201010_1ID	Cabezon de la Sal		ES	43.121	−4.292	Ha_59			unit31	**JQ921116**	**JQ921730**	**JQ921913**
P. strouhali (Karaman, 1955)	MARBRERI_200809_2ID	Perreon		FR	46.058	4.563	Ha_60			unit22	**JQ921427**	KC610114	KC610361
P. synaselloides (Henry, 1963)	THEOULE_200810_2ID	Méounes-lès-Montrieux		FR	43.264	5.960	Ha_61			unit20	**JQ921445**	**JQ921818**	**JQ921996**
P. vandeli Magniez & Henry, 1969	UTHURRIA_201005_2	Viodos-Abense-de-Bas		FR	43.234	−0.917	Ha_110			unit42	**JQ921490**	KC610158	KC610368
P. vulgaris (Sket, 1965)	DOVJEZ_200910_2ex	Dovjez		SI	46.114	14.485	Ha_111			unit12	**JQ921493**	**JQ921824**	**JQ922003**
P. walteri (Chappuis, 1948)	BURKH1_200905_1ex	Burkheim		DE	48.092	7.599	Ha_112			unit21	**JQ921498**	KC610159	×
Stenasellus breuili Racovitza, 1924	ALLI_200906_1ID	Larraun		ES	42.988	−1.893	Ha_115			unit6	**JQ921607**	**JQ921830**	×
Stenasellus racovitzai Razzauti, 1925	CASAMOZA_200810_1	Lucciana		FR	42.519	9.442	Ha_116			unit7	**JQ921621**	**JQ921836**	**JQ922011**

†Individuals present in the tree of Fig. 2.
‡ Hc_28 and Hc_44 included in Hc_16 and Hc_48 included in Hc_19 by TCS in the haplotype network for the *16S*.

Table S2 Priors used in BEAST for testing monophyly of *Proasellus valdensis* and presence of cryptic species.

Prior name	Prior definition	Value estimated	Method
ac	GTR A–C substitution parameter	Initial value of gamma [0.05, 10] prior distribution	Estimated by PHYML (Guindon *et al.*, 2010)
ag	GTR A–G substitution parameter	Initial value of gamma [0.05, 10] prior distribution	Estimated by PHYML
at	GTR A–T substitution parameter	Initial value of gamma [0.05, 10] prior distribution	Estimated by PHYML
cg	GTR C–G substitution parameter	Initial value of gamma [0.05, 10] prior distribution	Estimated by PHYML
gt	GTR G–T substitution parameter	Initial value of gamma [0.05, 10] prior distribution	Estimated by PHYML
Frequencies	Base frequencies		Default value
alpha	Gamma shape parameter	Initial value of uniform [0, 1000] prior distribution	Estimated by PHYML
pInv	Proportion of invariant sites parameter	Initial value of uniform [0, 1] prior distribution	Estimated by PHYML. Always equal to zero for the third codon position of *COI*, then set to 0.001
mu (*COI* only)	Relative rate parameter	Upper limit of uniform prior distribution	Estimated by the maximum difference between tip to root distance (calculated with APE; Paradis *et al.*, 2004) on PHYML phylogenies of each *COI* partition
tmrca	Time to most recent common ancestor statistic	Upper and lower limit of uniform prior distribution	Two calibration points were used: divergence between *Stenasellus breuilli* (Pyrenees) and *Stenasellus racovitzai* (Corsica) was set between 250 Ma (Neothetys opening; Stampfli, 2000; diversification of the Stenasellidae; Magniez, 2000) and 29 Ma (Corsica–continent separation; Orsini *et al.*, 1980). Divergence between *Galasellus heyli* (France) and *Caecidotea kenki* (Washington, DC) was constrained between 300 Ma (oldest isopod fossil discovered; Schram, 1970) and 54 Ma (North Atlantic opening; Skogseid *et al.*, 2000).
TreeModel.rootHeight	Root height of the tree	Initial value, mean and standard deviation of normal prior distribution	All set to 3, corresponding to the 300-Ma age of the oldest known isopod fossil (Schram, 1970)
yule.birthRate	Yule speciation process birth rate	Upper limit of uniform prior distribution	Set to the number of evolutionary units
meanRate	Mean rate of evolution over the whole tree	Upper limit of uniform prior distribution	Default value
coefficientOfVariation	Variation in rate of evolution over the whole tree	Upper limit of uniform prior distribution	Default value

Prior name	Prior definition	Value estimated	Method
covariance	Covariance in rates of evolution on each lineage with their ancestral lineages	Upper and lower limit of uniform prior distribution	Set to –100 and 100
ucld.mean	Uncorrelated lognormal relaxed clock mean		Default value
ucld.stdev	Uncorrelated lognormal relaxed clock standard deviation		Default value

REFERENCES

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

Magniez, G.J. (1999) A review of the family Stenasellidae (Isopoda, Asellota, Aselloidea) of underground waters. *Crustaceana*, **72**, 837–848.

Morvan, C., Malard, F., Paradis, E., Lefébure, T., Konecny-Dupré, L. & Douady, C.J. (2013) Timetree of Aselloidea reveals species diversification dynamics in groundwater. *Systematic Biology*, **62**, 512–522.

Orsini, J., Coulon, C. & Cocozza, T. (1980) Dérive cénozoique de la Corse et de la Sardaigne et ses marqueurs géologiques. *Geologie en Mijnbouw*, **59**, 385–396.

Paradis, E., Claude, J. & Strimmer, K. (2004) APE : analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Schram, F.R. (1970) Isopods from the Pennsylvanian of Illinois. *Science*, **169**, 854–855.

Skogseid, J., Planke, S., Faleide, J.I., Pedersen, T., Eldholm, O. & Neverdal, F. (2000) Ne atlantic continental rifting and volcanic margin formation. *Special Publications of the Geological Society, London*, **167**, 295–326.

Stampfli, G.M. (2000) Tethyan oceans. *Special Publications of the Geological Society, London*, **173**, 1–23.

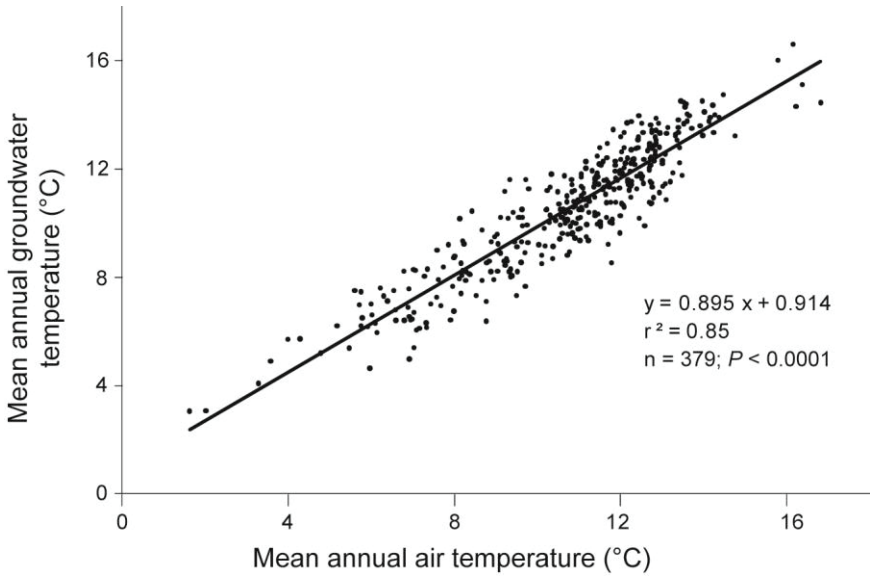
Journal of Biogeography

SUPPORTING INFORMATION

Integrating phylogeography, physiology, and habitat modelling to explore species range determinants

David Eme, Florian Malard, Céline Colson–Proch, Pauline Jean, Sébastien Calvignac, Lara Konecny–Dupré, Frédéric Hervant, and Christophe J. Douady

Appendix S2 Relationship between mean annual air temperature and mean annual groundwater temperature in the Alps and Jura Mountains, France.



Linear relationship between mean annual air temperature and mean annual groundwater temperature for 379 sites located in the study area. Groundwater temperature data were obtained from the ADES data base (ADES, 2009).

REFERENCE:

ADES, (2009). Portail national d'Accès aux Données sur les Eaux Souterraines. Available at: http://www.ades.eaufrance.fr

Journal of Biogeography

SUPPORTING INFORMATION

Integrating phylogeography, physiology, and habitat modelling to explore species range determinants

David Eme, Florian Malard, Céline Colson–Proch, Pauline Jean, Sébastien Calvignac, Lara Konecny–Dupré, Frédéric Hervant, and Christophe J. Douady

Appendix S3 Most likely topology of 69 Aselloidea morphospecies and *16S* statistical parsimony haplotype network for *Proasellus valdensis*.

Figure S1 Most likely topology of 69 Aselloidea species in Europe inferred under a GTR+G+I model of substitution. Genera of Aselloidea and main clades within *Proasellus* are shown on the right. Supports for these clades are shown along branches. The 69 morphospecies were selected to test for the monophyly of *Proasellus valdensis* in the Alps and Jura Mountains. Consequently, the tree is not intended to provide a fully resolved phylogenetic framework of the Aselloidea (see Morvan *et al.,* 2013 for further details).

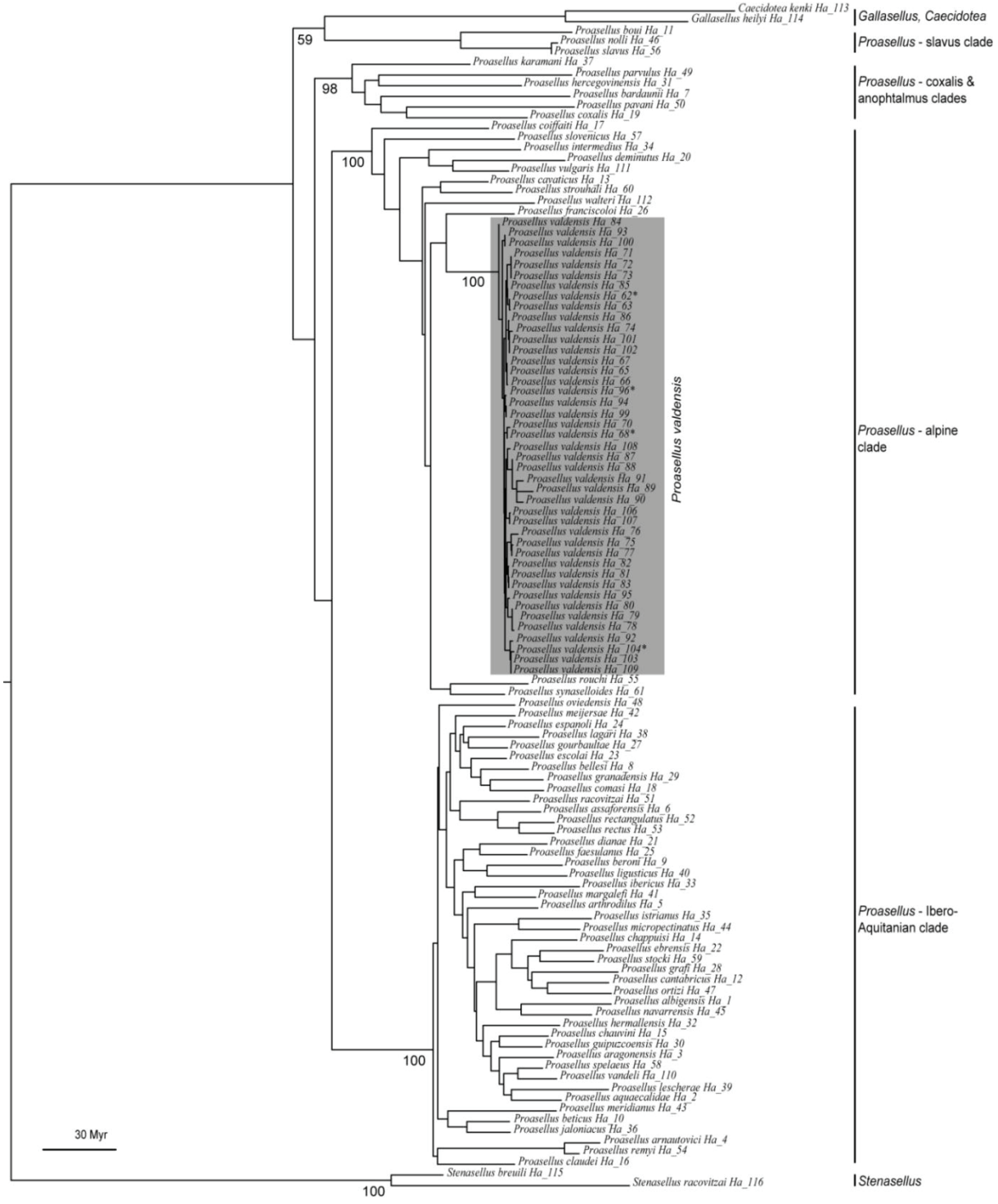
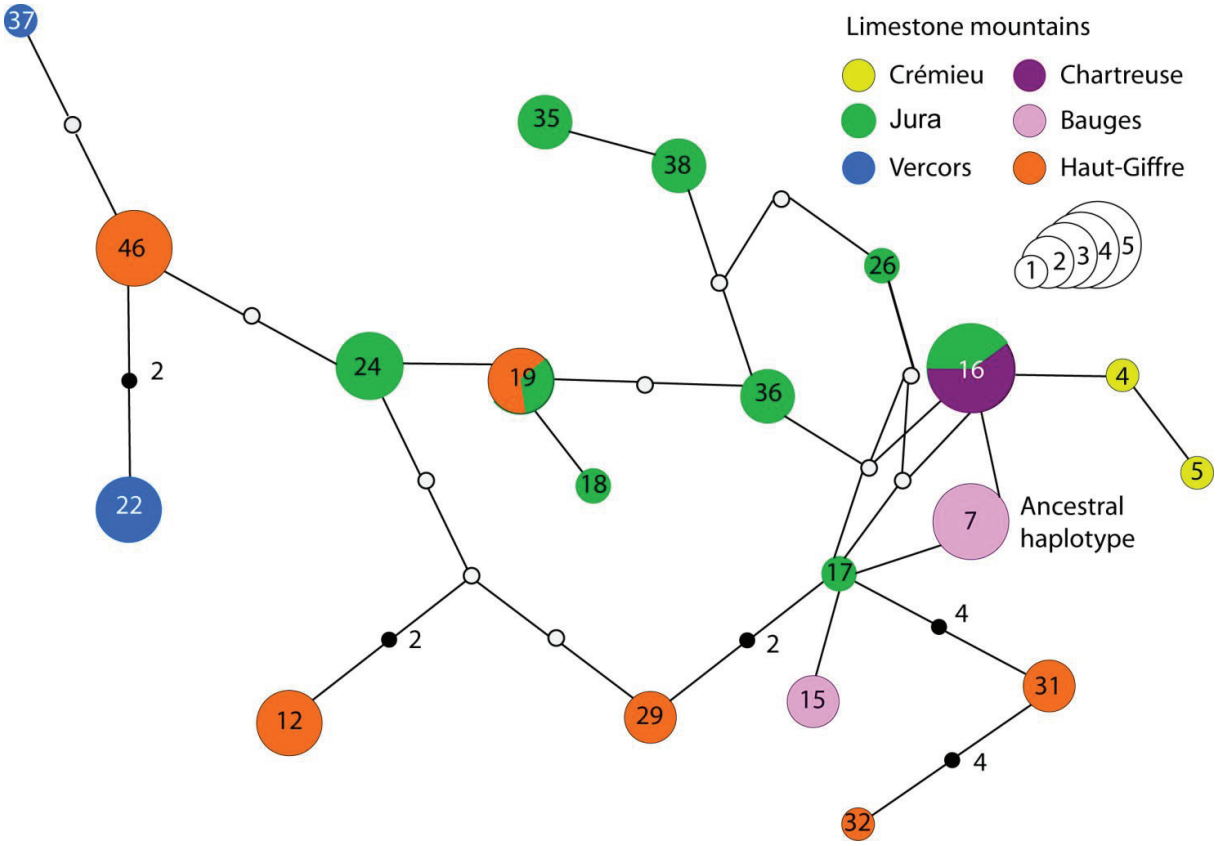


Figure S2 *16S* statistical parsimony network for *Proasellus valdensis* in the Alps and Jura Mountains (*n* = 44 individuals). The size of each colored circles is proportional to the number of

times that haplotype was sampled. Numbers within circles refer to haplotype codes (see Table S1 in Appendix S1). The size of the colored sectors within a pie chart is proportional to the numbers of individuals. White and black circles indicate non-sampled or extinct haplotypes. Numbers next to black circles indicate the consecutive number of non-sampled or extinct haplotypes.



REFERENCE:

Morvan, C., Malard, F., Paradis, E., Lefébure, T., Konecny-Dupré, L. & Christophe J. Douady, C.J. (2013) Timetree of Aselloidea reveals species diversification dynamics in groundwater. *Systematic Biology*, **62**, 512-522.

Annexe 7 : Article 6 : Independent, overlapping and interacting effects of human land use and geo-climatic factors on European patterns of diversity in freshwater ecosystems

Cette annexe présente un sixième article auquel j'ai participé en tant que co-auteur au sein d'un groupe de travail du programme européen BioFresh. Cet article, actuellement soumis à la revue *Freshwater Biology*, a évalué dans un cadre multi-groupes (« poissons », mollusques, macrophytes, invertébrés benthiques, et crustacés souterrains) et multi-écosystèmes (lotique, lentique et souterrain) l'importance relative des pressions anthropiques exercées par l'utilisation du paysage et des facteurs naturels sur les patrons de la biodiversité des eaux douces en Europe. J'ai participé à la conception de l'article et j'ai réalisé les analyses propres aux crustacés aquatiques souterrains. Les résultats mettent en avant le rôle dominant joué par les facteurs naturels sur les patrons de diversité pour l'ensemble des groupes taxonomiques et les différents écosystèmes étudiés. L'effet de l'utilisation des terres par l'homme (terres agricoles et urbaines) co-varie avec les variables environnementales naturelles. Dès lors, la part des effets attribuables uniquement aux facteurs anthropiques reste très faible, sans qu'il soit possible de détecter nécessairement un impact négatif fort sur la diversité biologique des milieux.

Title Independent, overlapping and interacting effects of human land use and geo-climatic factors on European patterns of diversity in freshwater ecosystems

Authors

Christian K. Feld 1‡, Sebastian Birk 1, David Eme 2, Michael Gerisch 3, Daniel Hering 1,

Martin Kernan 4, Kairi Maileht 5, Ute Mischke 6, Ingmar Ott 5, Florian Pletterbauer 7, Sandra

8 Poikane 8, Jorge Salgado 4, Carl Sayer 4, Jeroen van Wichelen 9 and Florian Malard 2

Affiliations

1 Faculty of Biology, Department of Aquatic Ecology and Centre for Water and Environmental Research (ZWU), University of Duisburg-Essen, Universit.tsstr. 5, 45141 Essen, Germany

2 Universit. de Lyon ; UMR5023 Ecologie des hydrosyst.mes Naturels et Anthropis.s ; Universit. Lyon1 ; ENTPE ; CNRS ; 6 rue Rapha.l Dubois, 69622 Villeurbanne, France.

3 Department of Conservation Biology, Helmholtz Centre for Environmental Research – UFZ, Permoserstr. 15, 04318 Leipzig, Germany

4 University College London, Gower Street, London WC1E 6BT, UK

5 Institute of Agricultural and Environmental Sciences, Centre for Limnology, Estonian University of Life Sciences, Rannu Parish, Tartu 61117 Estonia

6 Department of Ecohydrology, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 310, Berlin 12587 Germany

7 Institute of Hydrobiology and Aquatic Ecosystem Management, University 23 of Natural Resources and Life Sciences Vienna, Max Emanuel Stra.e 17, 1180 Vienna, Austria

8 European Commission, Joint Research Centre, Institute for Environment and Sustainability, via E. Fermi 2749, Ispra, VA I-21027 Italy

9 Protistology and Aquatic Ecology, Ghent University, Krijgslaan 281 – S8, Gent 9000, Belgium

Running head

Effects of land use on European freshwater biodiversity

Keywords

Freshwater ecosystems, biodiversity, arable land, urban area, geo-climatic descriptors, variance partitioning

Correspondence:

Christian K. Feld, Department of Aquatic Ecology, Faculty of Biology and Centre for Water and Environmental Research (ZWU), Universit.tsstr. 5, 45141 Essen, Germany; E-mail: christian.feld@uni-due.de

Abstract

1. Human life is highly dependent on fresh water, which results in high population densities, intensive land and water uses and modification and pollution hotspots in the vicinity of freshwater bodies. In particular, land use is considered one of the main stressors on biodiversity patterns of freshwater ecosystems, with up to 80% non natural regional land cover in Europe. Consequently human impacts on freshwater biodiversity are numerous and wide-ranging.
2. Here, we address the impact of arable and urban landscapes, on the diversity of 11organism groups encompassing vertebrates, invertebrates and plants, occurring in five freshwater ecosystems: rivers, floodplains, lakes, ponds and groundwater. In addition, nine geo-climatic variables (e.g. latitude, longitude, precipitation) were used to quantify the independent, overlapping and interacting effects of land use and natural descriptor variables. Biodiversity response was computed as taxon richness, Shannon diversity, taxon rareness and taxonomic distinctness.
3. The four biodiversity metrics were analysed using a variance partitioning scheme based on boosted regression trees (BRT) and subsequently with generalised linear modelling (GLM). The analyses sought: i) to partition the unique, shared and unexplained variation in the metrics explained by both groups of descriptor variables and ii) to quantify the contribution of each descriptor variable to biodiversity variation in the data.
4. Variance partitioning revealed the variation in biodiversity uniquely described by land use was consistently low across all ecosystems and organism groups. In contrast, the variation accounted for by both unique geo-climatic descriptors and the joint effects of both descriptor groups explained significantly more variance in the 39 biodiversity metrics tested. The GLM confirmed this and revealed significant interactions between geo-climatic descriptors and land use for roughly a third of the 66 33 GLM models. The interactions accounted for up to 17% of model deviance. With both BRT and GLM, however, no consistent patterns were observed related to the type of biodiversity metric and organism group considered.
5. Dividing the data according to the strongest geo-climatic gradient in each dataset was undertaken to reduce the strength of the respective natural descriptor variable and determine whether land use effects on biodiversity would increase in the data subsets. Results showed that data sub-setting can highlight land use effects on freshwater biodiversity, if geo-climatically more homogeneous datasets are analysed. However, the increased role of land use was not linked to the latitudinal or longitudinal extent of the data subsets, suggesting that the observed land use effects were not dependent upon the spatial extent of the subsets.
6. Our results confirm there are significant joint effects of, and interactions between, land use and natural environmental factors on freshwater biodiversity. This has three implications for biodiversity monitoring and assessment schemes. First, the combined analysis of anthropogenic stressors and geo-climatic factors is a prerequisite for the detection and

quantification of human threats to biodiversity. Second, geo-climatically more homogeneous datasets can unmask the role of anthropogenic stressor variables in the analysis. And third, whole community-based biodiversity metrics reveal contrasting response directions and thus should be complemented by other metrics which account for taxon identity and turnover, to better address the loss of biodiversity in response to land use impacts and other stressors.

Introduction

Although freshwaters cover only 1% of the earth's surface, almost 10% of the world's species live in freshwater ecosystems (Loh & Wackernagel, 2004). Freshwater biodiversity is declining faster than marine and terrestrial biodiversity (Dudgeon *et al.,* 2006), most likely because human life and many human activities rely on fresh water. This results in high population densities, intense land and water uses and modification and pollution hotspots in the vicinity of freshwater bodies. Consequently human impacts on freshwater biodiversity are numerous and wide ranging. Dudgeon et al. (2006) identify five major stressors of biodiversity which affect different freshwater ecosystem types to varying degrees: water i) overexploitation; ii) water pollution; iii) flow modification; iv) habitat degradation; and v) invasive species. While rivers are more affected by physical alterations (e.g. dams, impoundments, disconnection from the floodplain), lentic waters are more susceptible to nutrient enrichment (Wetzel *et al.,* 2001; Schindler, 2006), with increasing adverse effects on lentic biota under climate change (Jeppesen *et al.,* 2010; 2012).

Many of these stressors can be closely linked to land use, which may therefore be considered a composite (or proxy) stressor. Intensive agriculture, in particular, affects both lotic and lentic biodiversity through flow modification, pollution by fine sediment and pesticide fluxes (Allan, 2004; Feld, 2013), habitat degradation and eutrophication (Jeppesen *et al.,* 2000). Urbanisation represents another intensive land use, with strong effects on freshwater biodiversity, resulting in "consistent declines in the richness of algal, invertebrate, and fish communities" (Paul & Meyer, 2001). In Europe, a very high share (up to 80%) of the land is intensively used for settlements, infrastructure and production systems (including agriculture and intense forestry (http://www.eea.europa.eu/themes/landuse/intro; accessed on 1 April 2014) and aquatic biodiversity is impoverished accordingly. Although point source pollution caused by intensive land use has decreased in recent decades due to enhanced waste water treatment, the legacy effects on biodiversity may be long-lasting, representing 114 "the ghost of land use past" (Harding *et al.,* 1998).

Anthropogenic stress intensity and thus its influence on biodiversity differs regionally, impacting large-scale biodiversity patterns, originally shaped by natural drivers. These natural drivers are considered in macro-ecological studies focusing on i) landscape energy/climate, ii) area/habitat heterogeneity and iii) history (e.g. Mittelbach *et al.,* 2007; Leprieur *et al.,* 2011; Oberdorff *et al.,* 2011). The influence that landscape energy and climate have on biodiversity are primarily driven by temperature, precipitation or evapo-transpiration, all of which influence ecosystem energy supply and thus control or support biophysical processes operating within the system (Wright 1983; Hawkins *et al.,* 2003; Evans *et al.,* 2005; Mittelbach *et al.,* 2007; Field *et al.,* 2009). Area/habitat heterogeneity refers to the size and heterogeneity (habitat diversity) of an area under consideration, with the assumption that larger and more heterogeneous areas exhibit higher biodiversity (sensu Mc Arthur & Wilson, 1963; Gu.gan *et al.,* 1998; Davies *et al.,* 2007). Lastly, historical events (i.e. previous and often long-term events dating back for centuries or even millennia) may continue to shape

contemporary biodiversity patterns (Mittelbach *et al.,* 2007; Leprieur *et al.,* 2011; Tisseul *et al.,* 2012). The expansion of Pleistocene glaciers and their subsequent contraction followed by recolonisation, for example, are considered a key factor in explaining much of the variation in the distribution of contemporary biodiversity across Europe (Reyjol *et al.,* 2007; Araujo *et al.,* 2008; Baselga *et al.,* 2012), with formerly glaciated regions (e.g. Scandinavia) generally exhibiting less diversity than non-glaciated regions (e.g. Mediterranean peninsula). Over more recent timescales land use practices dating back decades may continue to shape contemporary biodiversity even if land use has subsequently changed or been abandoned (Harding *et al.,* 1998).

Both the natural drivers of freshwater biodiversity and multiple stressors 138 resulting from human land and water uses have been addressed in a multitude of studies (see Stendera *et al.,* 2012 for a recent summary of 368 papers), although few have considered these in an integrated way. Studies that investigate the combined effects of natural and anthropogenic descriptors are rare. Furthermore, Stendera et al. (2012) found that the majority of studies on natural drivers were rather broad-scale (continental and global), whereas studies on anthropogenic factors tend to focus on much finer (regional and local) spatial scales. The spatial resolution (grain size) also often differs, with the catchment 'grain' prominent in broad-scale studies, but single sites within one or several catchments foremost in fine-scale studies. Few studies addressed the impacts of both natural drivers and anthropogenic stressors on biodiversity and there remains a limited understanding of the synergies between both factors.

Brucet et al. (2013) suggest anthropogenic stressors have a minor role in shaping biodiversity patterns of lake fish assemblages in Europe compared to broad-scale climatic drivers. They found, for example, that the eutrophication gradient in their data was less significant than the natural temperature gradient. At the European scale however, these gradients (or drivers) are linked; eutrophication often results from intensive agriculture, the location of which is largely determined by recent and historic geo-climatic factors (e.g. altitude, mean annual temperature, annual precipitation, glaciation). As a result, intense row-crop agricultures (e.g. maize, rye, wheat) primarily occur in the temperate lowland regions of Central Europe (http://www.eea.europa.eu/data-and-maps/explore-interactive-maps/changing-face-of-europe-2014; accessed on 1 April 2014). Further, urbanisation, the second main composite stressor affecting freshwater biodiversity, is not independent from natural gradients; many large metropolitan areas in Europe (e.g. London, Paris, Cologne, Ruhr Metropolitan area, Berlin, Warsaw) are located between 50Åã and 52Åã N in lowland areas, i.e. within a narrow band of temperate climate conditions. Therefore, we expect strong interactions between land use and geo-climatic drivers and their impacts on freshwater biodiversity patterns. 164 Both factors may interact in different ways: agriculture is least intensive in Scandinavia, where biodiversity is low due to the legacy of glaciation; urbanisation is strongest in Central Europe, away from the extremes of temperature and altitude. Intensive agriculture is most prominent in Central Europe and the Mediterranean region, yet the Mediterranean region in particular was not affected by Pleistocene glaciers and thus is one of the key biodiversity hotspots worldwide (Myers *et al.,* 2000).

In this study, we developed a stepwise analysis to determine the independent, overlapping and interacting effects of anthropogenic land use and geo-climatic factors on the European biodiversity patterns of eleven organism groups in five lentic and lotic ecosystem types (rivers, lakes, floodplains, ponds and groundwater). First, we used a machine-learning technique to partition the variance and to quantify the independent and overlapping effects of both factors in each ecosystem. Second, we performed regression modelling including interactions of both factors and tested the significance of interaction terms. Eventually, to decrease the effect of the most influential geo-climatic variable in the regression models, we generated subsets of the data and quantified the proportion of variance

attributable to land use separately for each subset. This is the first study to address the unique, shared and interacting effects of geo-climatic variables and land use on freshwater biodiversity patterns across numerous ecosystem types and organism groups.

Methods

Stressor variables

For all but groundwater ecosystems we used CORINE land cover data (European Environmental Agency; http://www.eea.europa.eu/publications/COR0-landcover) to calculate the proportion of arable and urbanised land within a catchment or the area directly surrounding a site (Table 1). The area considered differed between ecosystem types and was selected to match the scale of biological sampling. The CORINE land cover data are based on satellite imagery (Landsat 7, 25 x 25 m pixels), cover most countries in Europe (geometric accuracy: 100 m) and encompass land cover types with a minimum area of 25 ha. We used the land cover classes 'arable land' and 'urban land' (hereafter referred to as land use), which aggregate the CORINE level 3 types '2.1.1 Non-irrigated arable land' as 'arable land' and the level 2 types '1.1 Urban fabric' and '1.2 Industrial, commercial and transport units' as 'urban land'. We focused on these two land use types, because they are known to strongly affect aquatic biodiversity via numerous individual stressors (Paul & Meyer, 2001; Allan, 2004; Feld, 2013). For groundwater systems, we used the GlobCover land cover data (http://due.esrin.esa.int/globcover/) due to its comprehensive coverage of Eastern Europe. 'GlobCover Land Cover v2' is a global land cover map at a resolution of 10 arc seconds (or 300 m at the equator) and corresponds well with the CORINE land cover classification. Arable and urban land uses were derived from a grid-based scheme throughout Europe, with a grid size of 100 x 100 km (EDIT geoplatform; Sastre *et al.,* 2009). The same grid was applied to generate the land use data for lakes using the CORINE land cover data. Proportions of different land use types were obtained by clipping the land use maps (either CORINE or GlobCover) with a layer containing the polygonal information from the targeted areas (Table 1) within a geographic information system (ESRI ArcGIS 10, Redlands, CA).

Table 1: Spatial scale considered and data sources used to generate arable and urban land use data.

	Land use data source	Area [km^2]	Shape of area	Comment
Rivers	CORINE 2006	variable	irregular	entire catchment upstream of a site
Floodplains	CORINE 2006	78.5 km^2	circle around site (radius = 5 km)	
Ponds	CORINE 2006	4.9 km^2	circle around pond (radius = 125 m)	
Lakes	CORINE 2006	10,000 km^2	100 x 100 km grid	
Groundwater	GlobCover	10,000 km^2	100 x 100 km grid	

Geo-climatic descriptor variables

We used nine natural environmental descriptors covering geographical and climatic variables (hereafter referred to as geo-climatic variables, Supplementary Table S1). Latitude, longitude, altitude and catchment size were derived from digital maps using ArcGIS 10. Latitude and longitude were included as proxy geographical variables representing other potential natural drivers of biodiversity, such as historical climate and glaciation (Hortal *et al.,* 2011; Stendera *et al.,* 2012), but were excluded from the analysis if they were collinear with any of the other environmental descriptors. Altitude was included to account for the role of topography in shaping diversity patterns (e.g. Davies *et al.,* 2006). Lake surface area was derived from the WISER lake database (Moe *et al.,* 2013). Mean annual air temperature and annual precipitation were abstracted from the WorldClim database version 1.4 (Hijmans *et al.,* 2005). WorldClim summarises measured data at weather stations between 1950 and 2000 as monthly mean values, interpolated by a thin-plate smoothing spline algorithm to fit a raster grid (grid size: 30 arc seconds, approximately 1 km at the equator). Mean annual air temperature was averaged from long-term yearly means, whereas a yearly mean was averaged from monthly means throughout a year. Annual precipitation was based on the sum of long-term monthly mean precipitation values. Actual and potential evapo-transpiration (AET, PET) were derived from the CGIAR-CSI Global-PET database (for details, see Zomer *et al.,* 2008; http://www.cgiar-csi.org).

Biological data

Rivers

Site-specific river data were derived from the WISER river database (Moe *et al.,* 2013), encompassing taxa lists of fish, macroinvertebrate and macrophyte communities and proportional catchment land use for up to 1,221 sites across Central Europe (Central/Western Mountains and Central/Western Plains ecoregions of France, Germany and 235 Austria; Illies, 1978). Macroinvertebrate data were available for all sites, fish data for 590 sites and macrophyte data for sites. The taxa lists originate from national monitoring surveys and followed the national monitoring standards defined for field sampling methodology and sample processing (see Dahm *et al.,* 2012 and Feld, 2013 for details).

Prior to the calculation of biodiversity metrics, the raw taxa lists obtained from the WISER river database were manually adjusted to eliminate researcher-dependent bias, for example, caused by different taxonomic determination levels for macroinvertebrates (e.g. Oligochaeta, Diptera). Species-level identification was achieved for fish and macrophytes, while genus level was used for macroinvertebrates, as this is the standard determination level in France.

Lakes

Lake phytoplankton taxa lists from 836 lakes (surface area >0.5 km2) in 20 European countries were derived from the WISER lake database (Moe *et al.,* 2013). The lakes are distributed among three major European regions: i) the Mediterranean region (145 lakes in Cyprus, Italy, Spain, Portugal and Romania), ii) the Central/Baltic region (373 lakes in Belgium, Germany, Estonia, France, Hungary, Lithuania, Latvia, the Netherlands and Poland) and iii) the Northern region (318 lakes in Denmark, Finland, Ireland, Norway, Sweden and the United Kingdom).

We chose samples taken between 2004 and 2010 to maximise the temporal comparability of samples. If multiple samples were available for a lake within this period, we selected the most recent

sampling occasion to avoid a sample-density bias. For each sample, all stations within the same water body were combined by averaging to create a mean abundance for each lake. Taxa records from each country were harmonised for nomenclature (Phillips *et al.,* 2012).

Ponds

We defined ponds as shallow lentic water bodies with surface area less 259 than five hectares (0.05 km2) (De Meester *et al.,* 2005). Pond taxa lists were obtained from 32 peer-reviewed publications indexed in the Web of Science and generated for amphibians, macrophytes and macroinvertebrates (Gastropoda, Odonata and Coleoptera only). Additional data were collated from Homes, Hering & Reich (1999), Nagorskaya et al. (2002), Sobkowiak (2003), Oertli et al. (2005), Sayer, Davidson & Jones (2010), B.hmer (2012), Moe et al. (2013), the European Pond Conservation Network (http://campus.hesge.ch/epcn), N.J. Willby (University of Stirling, UK; unpubl.) and B.A. Luk.cs (Hungarian Academy of Sciences, HU; unpubl.). Amphibian species were recorded once at 148 ponds in seven European countries.

Macrophyte species records comprised 601 samples at 392 ponds in eight countries (genus level targeted for Chara sp. and Callitriche sp.; only hygrophytes, helophytes and hydrophytes with Ellenberg's moisture values ≥ 7 and stoneworts considered; Ellenberg *et al.,* 1992). Macroinvertebrate taxa lists were collated using 189 samples from 176 ponds in twelve countries (species or genus level). Due to heterogeneous and thus incomparable sampling efforts, only binary data (i.e. presence/absence) were generated.

Floodplains

The floodplain database is based on publications of European datasets on plants, ground beetles and molluscs in riverine wetland ecosystems. A literature review was conducted using Web of Science, covering publications between 1990 and 2012. Altogether, 78 publications were reviewed in detail to generate three taxa lists (total number of sample sites: 565): 352 sites for floodplain vegetation, 132 sites for ground beetles and 81 sites for molluscs. Samples from distinct and separated habitat types within the same floodplain counted as different sites. The sites are located in 21 countries and on 51 river floodplains across Europe, with the majority of sites located in Central Europe: Poland: 99 sites, Germany: 98, France: 81, Belgium: 42, Switzerland: 29, the Netherlands: 25, Czech Republic: 7 and Denmark: 6. Standardisation of species abundances among studies was impossible 284 due to the lack of information on sampling effort in most studies. The bias in sampling effort was minimised by omitting studies with an extremely short or long field sampling period and those with strongly skewed or otherwise inconsistent data.

Groundwater

The European groundwater crustacean data set (EGCD) was assembled as part of the European BioFresh project (http://data.freshwaterbiodiversity.eu/metadb/bf_mdb_view.php?uid=5326d79b4af7b&code=60). It covers the whole of Europe, except Russia, and contains a total of 21,700 database records, which collectively represent 12 orders and 1,570 species and subspecies of obligate groundwater Crustacea. Records are from the European PASCALIS database (Deharveng *et al.,* 2009), the hypogean crustacean recording scheme United Kingdom (Knight, 2012), the distributional checklist of the Italian fauna (Ruffo & Stoch, 2006), and the Berlin museum collection. They were complemented with occurrence data from an extensive literature search (i.e. 1,380 literature

sources representing half of the records in the EGCD). Species names and distributions were checked by taxonomic experts and spurious occurrences were excluded from the data set. Occurrence data were projected onto the grid of 0.9° latitude cells provided by the EDIT geoplatform (Sastre *et al.,* 2009). The area of cells in the grid was kept constant (10,000 km2) by adjusting the longitudinal divisions between adjacent cells in each latitudinal band. The final grid had 701 cells, 494 of which contained at least one species occurrence.

Calculation of biodiversity metrics

Biodiversity has many facets and, amongst others, encompasses compositional (structural), functional (trait) and phylogenetic aspects of assemblages. Given the mixture of binary (presence/absence) and continuous (abundance) data, the set of biodiversity metrics commonly calculable across all ecosystems was restricted here to total 309 species richness, species rareness and taxonomic distinctness (i.e. phylogenetic diversity). With abundance data, we also calculated Shannon-Wiener diversity (referred to as Shannon diversity in the following). Species richness and Shannon diversity are among the most commonly-used indicators of aquatic biodiversity in Europe (see Birk *et al.,* 2012 for a recent review of monitoring methodology). Taxon rareness (or endemicity) can be derived using the index of endemicity proposed by Crisp et al. (2001) and Linder (2001). The index describes the sum of relative frequencies of all taxa encountered at a site or within an area (grid) in relation to the overall number of sites or areas (grids) where the individual taxa have been observed. Hence, the index provides a measure of the summed relative frequencies of 'endemic' (or rare) taxa within a community, based on the overall frequency of the taxa in the entire dataset. Taxonomic distinctness refers to the mean taxonomic dissimilarity of any pair of taxa within a community along a Linnean phylogenetic tree (species, genus, family, order, class, phylum; Clarke & Warwick, 1998; 1999). For example, three species of the same genus are taxonomically less distinct than three species of different genera, orders or higher taxonomic entities, which is why taxonomic distinctness is also referred to as phylogenetic diversity. Taxonomic distinctness is applicable to binary taxa lists and adds a unique aspect of biodiversity, neither covered by taxon richness nor by taxon evenness (Gallardo *et al.,* 2011; Feld *et al.,* 2013).

Data analysis

We applied a stepwise analytical protocol for the multivariate analysis using Boosted Regression Tree analysis (BRT) and Generalised Linear Modelling (GLM).

STEP 1

Individual BRTs were run for each possible combination of organism group 333 and biodiversity metric using all geo-climatic and land use descriptors (full model) to compare the effects of both descriptor groups. The major advantages of BRT analysis over classical regression modelling are its capacity to i) analyse collinear descriptor variables, ii) handle non-linear descriptors with missing values and iii) identify interactions between descriptors (Elith, Leathwick & Hastie, 2008).

The full BRT models allowed us to identify the contribution of each individual descriptor' to the overall variance explained in a biodiversity metric and the pairwise interactions between descriptor variables. Both were then used in GLM (see second step) to define the entry order of each descriptor variable in a model and the interaction terms (see below). Partial Dependence Plots (PDP) enabled the

identification of the response patterns of biodiversity metrics along environmental descriptor gradients (Cutler *et al.,* 2007). Accordingly, the PDPs identified potential thresholds along the geo-climatic gradients at which a biodiversity metric value either sharply increased or decreased (Clapcott *et al.,* 2012; Feld, 2013a). Such thresholds may mark natural split points in the data, for example, geographical splits at a specific latitude, longitude or altitude, which then imply the presence of spatial patterns in the targeted biodiversity metric. We subsequently used these split points for the strongest geoclimatic descriptor in each BRT to divide each dataset (i.e. ecosystem type x organism group) into two subsets (see third step).

In addition to the full BRT models, we applied an additive partial regression scheme following Legendre & Legendre (1998, p. 531) to decompose the explained variation of the biodiversity metrics into four fractions: i) pure geo-climatic, ii) pure land use, iii) shared geoclimatic/land use and iv) unexplained. The shared fraction (iii) represents the variation that may be attributed to geo-climatic and land use descriptors together and is obtained additively in partial regression. As such, it is inherently different from non-additive interaction terms as introduced into the GLM (see next step). Differences in the variance explained were tested for significance using a Wilcoxon signed rank test.

STEP 2

GLM was applied individually to each combination of organism groups and three biodiversity metrics (Shannon's diversity excluded), and a set of geo-climatic and land use descriptors that excluded highly collinear variables, defined as those with a variance inflation factor >8 (Zuur, Ieno & Smith, 2007). We choose GLM for this step because of its flexibility in identifying the most parsimonious model (i.e. the best trade-off between model fit and complexity), including interactions between anthropogenic and geo-climatic descriptors. Adjusted goodness of fit (R2) and Akaike Information Criterion (AIC) were used as GLM quality indicators. The order of entry of each descriptor variable into a GLM model was based on the individual explanatory strength of the variable as identified in step 1 (i.e. the strongest descriptor entered a model first, followed by the second strongest, and so on). This procedure ensured a standardised and hence comparable analytical procedure for GLM models for all ecosystems. We used Poisson regression for species richness and Gaussian regression for rareness/endemicity and taxonomic distinctness. If overdispersion was detected in Poisson regression, we used negative binomial distribution functions in GLM. Rareness and taxonomic distinctness were logit-transformed to better-fit Gaussian regression (Warton & Hui, 2011). The GLM model with the highest explained deviance (equivalent to R2 in Gaussian GLM), in combination with the lowest AIC obtained for each combination of organism group and biodiversity metric, was selected as the final model. A final model included borderline significant descriptors ($0.05<P<0.1$), if the explained deviance and/or AIC notably improved with the descriptors in the model.

STEP 3

The final analytical step repeated the procedure for step 2, but was applied to 382 the data subsets. These subsets were defined using the split points of the most influential geo-climatic descriptor variable in each analysis. This was derived individually for each metric from the partial dependence plots of the BRTs (step 1). If necessary, the split points were slightly adjusted, to better achieve a balanced sample size of both data subsets. The objective of splitting the data according to the most influential geo-climatic descriptor variable was to control for the variance driven by the respective geo-climatic descriptor and thus to focus more on the role of land use.

All statistical analyses were run in R 2.15.3 (R Development Core Team, 2013). For BRTs, we

used the packages ‘gbm’ (Ridgeway, 2013) and ‘dismo’ (Hijmans *et al.,* 2013). GLMs were run with the package ‘MASS’ (Venables & Ripley, 2002).

Results

Partitioning the variance in biodiversity among geo-climatic and land use descriptors

Across all ecosystems, a total of 39 biodiversity metrics were calculated for eleven organism groups (Figure 1). Together, geo-climatic and land use descriptors explained between 20 and 93% (mean: 35%, SD: 18.7%) of the total variance in the full BRT models. On average, the explained variance was much higher for pond and floodplain biodiversity compared with the values obtained for the river, lake and groundwater models. No consistent metric driven differences across ecosystems and/or organism groups were detectable.

The variance partitioning scheme (BRT) revealed a consistently low proportion of variance attributable to pure land use effects for all metrics (Figure 1). Conversely, pure geo-climatic effects explained a relatively high proportion of the variance in river, lake and groundwater organisms and in pond amphibians and insects. This was irrespective of the biodiversity metric considered. Pure geo-climatic effects were significantly higher 406 than pure land use effects (Wilcoxon signed rank test: $P<0.001$), i.e. geo-climatic descriptors were significantly more influential than human land use for the observed biodiversity patterns.

However, the proportion of variance jointly attributable to both descriptor groups was equally high in many cases and particularly pronounced with the floodplain and pond results (Figure 1). It accounted for as much as 19–87% of the total variance in the floodplain biodiversity metrics (ponds: 35–63%). It was also comparatively high for rivers (0.6–41%), but much lower for lakes and groundwater (<12 and <10%, respectively for all metrics). Nevertheless, the joint effects of land use and geo-climatic variables were significantly higher than the effects of land use alone. The findings suggest that both descriptor groups were intrinsically allied in many models, which rendered the separation of its unique effects on the response variables difficult.

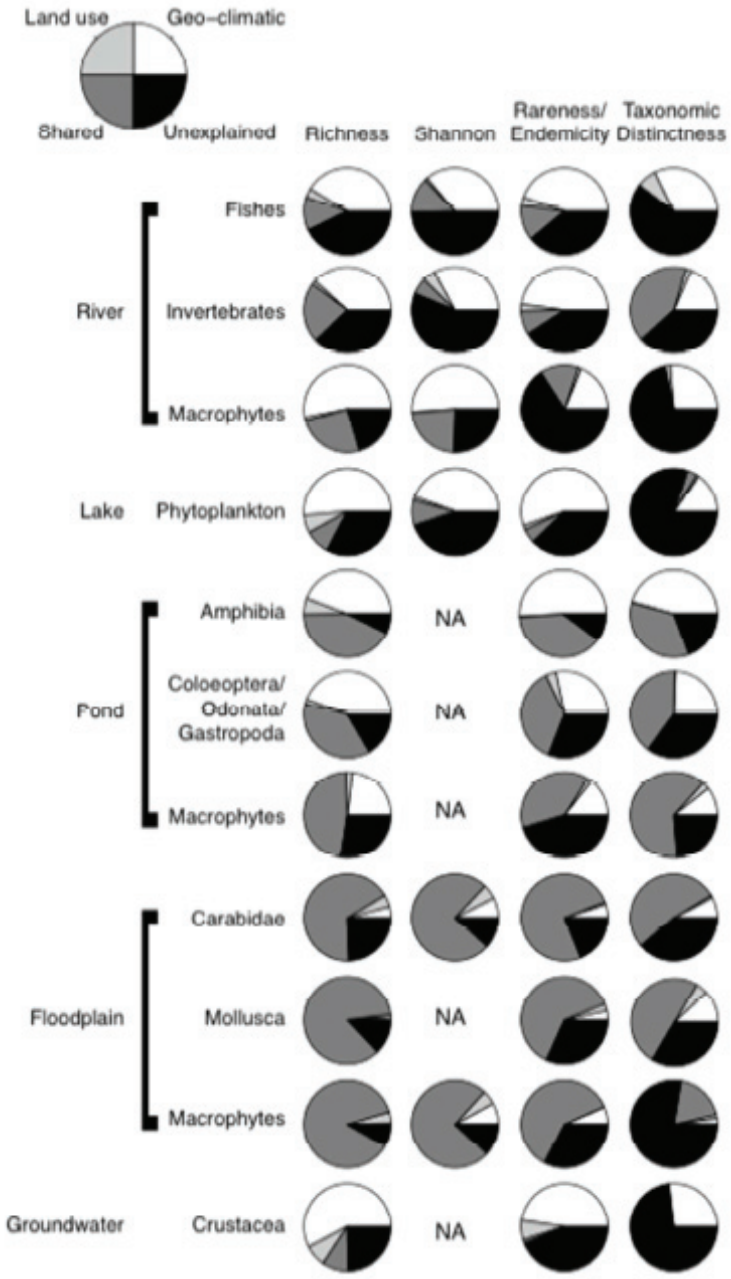


Figure 1: Variance partitioning scheme using four biodiversity metrics and eleven organism groups sampled in five ecosystem types. Each plot displays the pure and shared proportions of variance explained by land use and geo-climatic variables in the Boosted Regression Tree analyses (see text for details). NA = Shannon's diversity cannot be computed with presence/absence data.

Quantifying land use effects on biodiversity and interactions with geo-climatic descriptors

Similar as with the BRT results, land use descriptors alone accounted for less than 3% of the deviance (variation) in most GLM models (Table 2). Higher values (>10%) were found only for pond insect and floodplain carabid beetle richness and for river invertebrate and pond amphibian taxonomic distinctness. Both urban and agricultural land use performed similarly in the models and no general pattern was obvious regardless of the biodiversity metric considered.

Unexpectedly, however, we did not find a consistent decline in biodiversity in response to increasing land use intensity (Table 2). More often than not the sign of the relationship was positive, i.e. the biodiversity metrics value increased with increasing percentages of arable and urban areas. Irrespective of the biodiversity metric, organism group or ecosystem type, no consistent patterns were apparent.

Table 2: Matrix of strength and direction of biodiversity metrics in response to urban and agricultural land use across all ecosystem types and organism groups. Response strengths and direction ('+': positive, '–': negative relationship) are according to the highest deviance explained by land use (without interaction terms) in the GLM models using the complete datasets: >|10%| = +++/– – – ; >|5%| = ++/– –; >|3%| = +/–; ≤|3%| = O.

		Richness		Rareness/endemicity		Taxonomic distinctness	
Ecosystem	Organism group	Urban	Arable	Urban	Arable	Urban	Arable
Rivers	Fishes	O	+	O	++	O	O
Rivers	Invertebrates	O	–	O	O	++	+++
Rivers	Macrophytes	– –	O	O	– –	O	O
Lakes	Phytoplankton	O	O	O	O	O	O
Ponds	Amphibia	O	+	O	O	+++	– –
Ponds	Coleoptera/Odonata/Gastropoda	O	+++	++	O	O	O
Ponds	Macrophytes	O	++	O	++	–	O
Floodplains	Carabidae	– – –	– –	++	–	O	O
Floodplains	Mollusca	O	+	O	O	O	– –
Floodplains	Macrophytes	– –	O	O	O	O	O
Groundwater	Crustacea	+	O	O	O	O	– –

Significant interactions of geo-climatic descriptors and land use were found 431 for roughly a third of the 33 GLM models and accounted for up to 17% of model deviance (Table 3). The highest interactions (>10% explained deviance) were observed for floodplain carabid beetles and molluscs and for pond amphibians, but the majority of interaction terms accounted for less than were 5% of the deviance in the models. Land use interactions were strongest with longitude, latitude or annual precipitation, again highlighting the intrinsic co-dependence between land use and geo-climatic factors. Thus the land use patterns within these data were not independent of the geo-climatic patterns (or more specifically, the latitudinal and longitudinal location, respectively).

Table 3: Percent deviance explained by significant interaction terms including land use in the GLM models based on the complete datasets. If more than one interaction was significant, the total deviance explained by all interactions is provided. Geo-climatic descriptor(s) interacting with land use are listed in brackets; area = catchment size; lat = latitude; lon = longitude; ppt = annual precipitation; temp= mean annual air temperature; pet = potential evapotranspiration; hab = habitat diversity.

		Richness		Rareness/endemicity		Taxonomic distinctness	
Ecosystem	Organism group	Urban	Arable	Urban	Arable	Urban	Arable
Rivers	Fishes		1.4 (area)				
Rivers	Invertebrates						
Rivers	Macrophytes				1.5 (lat)		
Lakes	Phytoplankton						
Ponds	Amphibia	14.4 (lon)			3.1 (lon)	11.3 (lon)	
Ponds	Coleoptera/Odonata/Gastropoda			5.1 (lat, ppt, temp)	6.6 (lat)	2.7 (ppt)	
Ponds	Macrophytes	1.8 (ppt)	2.3 (lat, ppt)	3.7 (pet)			1.2 (lat)
Floodplains	Carabidae		11.1 (ppt)	9.5 (ppt)			

		Richness		Rareness/endemicity		Taxonomic distinctness	
Ecosystem	Organism group	Urban	Arable	Urban	Arable	Urban	Arable
Floodplains	Mollusca	4.1 (lon)		17 (ppt)			
Floodplains	Macrophytes				3.7 (temp)		1.9 (lon)
Groundwater	Crustacea	1 (hab)					

Controlling the influence of geo-climatic descriptors by data sub-setting

Latitude or longitude explained a considerable fraction of the variation in many biodiversity metrics, regardless of the analytical approach applied. For example, in 11 out of 33 BRT models, either latitude or longitude was the strongest geo-climatic descriptor, followed by temperature (9 models), altitude (5), river catchment/lake surface area (3) and precipitation (3) (Table 4). Temperature and precipitation, however, are also linked to latitude and longitude at the European scale. By splitting the datasets along one of these (mostly) geographical gradients the intention was to reduce the geographical extent of the derived data subsets and hence would decrease the role of geo-climatic descriptors relative to the role of land use in the data subsets.

Indeed, our findings confirm that data sub-setting can control the analysis of land use effects on freshwater biodiversity, yet apparently not necessarily through a reduction in the spatial extent of the obtained data subsets. With floodplain mollusc richness, for example, the deviance explained by one subset (annual precipitation ≤630 mm, see Table 4) was five times the deviance explained by the full data and accounted for 50% of the metric's total deviance in this subset. Likewise, the respective values doubled with floodplain 456 carabid beetle and mollusc rareness/endemicity and achieved explained deviances between 40 and nearly 50% for one data subset (Figure 2, see Table 4 for the respective split points). In some cases, land use explained substantially more deviance in the biodiversity metrics in both subsets (e.g. groundwater crustacean richness and rareness/endemicity, Figure 2). However, all but one of these data subsets were obtained by splits along gradients of actual or potential evapotranspiration, mean annual air temperature or altitude (Table 4).

More generally, the changes observed in the deviance explained by land use (including interaction terms) when analysing the data subsets were largely independent of the changes in the geographical extent within the subsets (Figure 3). Neither latitudinal nor longitudinal splits of the full data resulted in consistent and significant increases (or decreases) in the deviance explained by the GLM models.

Table 4: Split points used to generate two data subsets for each combination of ecosystem type, organism group and biodiversity metric. Split points were identified using the partial dependence plots provided by the Boosted Regression Tree models, but were modified in order to achieve a more balanced sample size in both subsets. For clarity, subset 1 always encompasses the samples ≤ split point and subset 2 the samples > the split point. See text for details.

Ecosystem	Metric	Strongest geo-climatic descriptor variable	Split point	Sample size subset 1	Sample size subset 2
Rivers	Fish richness	Catchment size	500 km²	516	74
Rivers	Fish rareness	Catchment size	500 km²	516	74
Rivers	Fish taxonomic distinctness	Catchment size	500 km²	516	74
Rivers	Invertebrate richness	Latitude	51° N	639	582
Rivers	Invertebrate rareness	Latitude	51° N	639	582
Rivers	Invertebrate taxonomic distinctness	Latitude	51° N	639	582
Rivers	Macrophyte richness	Longitude	6° E	96	555
Rivers	Macrophyte rareness	Latitude	51° N	292	359
Rivers	Macrophyte taxonomic distinctness	Latitude	51° N	191	303
Lakes	Phytoplankton richness	Mean annual air temperature	6 °C	192	644
Lakes	Phytoplankton rareness	Mean annual air temperature	7.7 °C	315	521
Lakes	Phytoplankton taxonomic distinctness	Mean annual air temperature	9.3 °C	655	181
Ponds	Amphibia richness	Mean annual air temperature	8.8 °C	110	38

Ecosystem	Metric	Strongest geo-climatic descriptor variable	Split point	Sample size subset 1	Sample size subset 2
Ponds	Amphibia rareness	Ecoregion	4 (yes/no)	alpine: 84	non-alpine: 64
Ponds	Amphibia taxonomic distinctness	Mean annual air temperature	8.9 °C	89	35
Ponds	Coleoptera/Odonata/Gastropoda richness	Annual precipitation	992 mm	119	58
Ponds	Coleoptera/Odonata/Gastropoda rareness	Latitude	48 °N	109	62
Ponds	Coleoptera/Odonata/Gastropoda taxonomic distinctness	Mean annual air temperature	8.8 °C	53	124
Ponds	Macrophyte richness	Latitude	49 °N	338	263
Ponds	Macrophyte rareness	Latitude	49 °N	338	263
Ponds	Macrophyte taxonomic distinctness	Annual precipitation	839 mm	238	327
Floodplains	Carabidae richness	Annual mean air temperature	9.9 °C	62	70
Floodplains	Carabidae rareness	Altitude	37 m a.s.l.	62	70
Floodplains	Carabidae taxonomic distinctness	Altitude	55 m a.s.l.	76	55
Floodplains	Mollusca richness	Annual precipitation	630 mm	51	30
Floodplains	Mollusca rareness	Longitude	16.5 °E	32	47
Floodplains	Mollusca taxonomic distinctness	Longitude	12.3 °E	32	47

Ecosystem	Metric	Strongest geo-climatic descriptor variable	Split point	Sample size subset 1	Sample size subset 2
Floodplains	Macrophyte richness	Annual mean air temperature	9.9 °C	170	182
Floodplains	Macrophyte rareness	Altitude	49 m a.s.l.	150	202
Floodplains	Macrophyte taxonomic distinctness	Altitude	19 m a.s.l.	124	198
Groundwater	Crustacea richness	Evapotranspiration (AET)	600 mm	406	120
Groundwater	Crustacea endemicity	Mean annual air temperature	10.9 °C	134	256
Groundwater	Crustacea taxonomic distinctness	Altitude	462 m a.s.l.	217	121

Discussion

Pure and shared land use effects on freshwater biodiversity

Human land use, in particular urbanisation and intensified agriculture, are widely recognised as major threats to freshwater biodiversity worldwide (MEA, 2005; Dudgeon *et al.,* 2006; Vörösmarty *et al.,* 2010) and have been found to significantly impact the integrity of freshwater systems (e.g. Allan, 2004; Feld *et al.,* 2011, Feld, 2013). However, the variance partitioning scheme applied in this study to quantify the role of land use in comparison to the natural drivers of biodiversity reveals a consistently low proportion of variation in biodiversity explained purely by land use at the European scale. This is irrespective of the ecosystem type, organism group and biodiversity metric considered. In contrast, the natural geo-climatic descriptors are much better correlates of diversity. This suggests both land use variables are less influential compared to the geo-climatic gradients at the ecoregional and continental scales as addressed in this study and this is supported by the findings of Davies et al. (2006), who found that land use had weak explanatory power at the scale of biogeographic regions, but had a stronger role at the global scale.

More importantly, land use and geo-climatic variables exhibited strong shared effects, significantly higher than the pure land use effects. These shared effects imply a strong collinearity of both descriptor groups, which translates to highly concordant patterns of land use, geo-topographical and climatic conditions. It suggests that land use is not independent of geo-climate at the geographical scale covered by our data. This does not mean that land use effects on biodiversity are subordinate to geo-climatic drivers, but they simply cannot be fully disentangled and thus should be considered in tandem. In a similar study, Brucet et al. (2013) regressed fish diversity metrics in 1,632 European lakes against a selection of anthropogenic stressor variables and natural (geographic) descriptors. They

reported that 'geographical factors dominate over anthropogenic pressures', which is largely supported by our data, but also requires qualification in that geo-climatic factors not only dominate but act in concert with land use. This is important to distinguish and raises the issue of interaction between both descriptor groups (see next paragraph). As a consequence, studies that consider only one descriptor group run the risk of overlooking the strong shared explanatory power of land use and geo-climatic factors.

Interactions of land use with geo-climatic factors

The high proportion of shared variation revealed by the BRT analyses suggests an interaction of variables in both descriptor groups. This was further investigated and quantified by 33 GLM models, a third of which included significant interaction terms providing evidence for the combined effect of both descriptor groups. In particular latitude, longitude and annual precipitation most often interact with land use, reflecting a geographical and (historic) climatic pattern in the distribution of urban and agricultural areas in Europe. Since most interactions account for less than five percent of the model deviance and since significant interactions are not found in two thirds of models, we conclude that the additive shared effects obtained from BRTs cannot be translated to the multiplicative interactions identified by GLM. We are unable to explain further the nature of this linkage or interpret with any confidence the interaction of geo-climatic and anthropogenic gradients. Further investigation using the spatial distribution of biodiversity (i.e. the potential spatial pattern) in Geographic Weighted Regression (GWR) may help to locate regions where the shared effect of land use and geoclimatic factors is particularly strong (Gouveia *et al.,* 2013).

The role of geo-climatic descriptors in smaller data subsets

The dominant role of geo-climatic descriptors (altitude, latitude and longitude) over human impact at ecoregional or continental scales may be explained by the relatively short human impact gradients at both scales in comparison to climatic patterns (Davies *et al.,* 2006). We, therefore, hypothesised that data subsetting along the major geo-climatic descriptor gradients (i.e. cutting the gradient) would enhance the land use effects on biodiversity. Our results partly confirm the hypothesis, but generally

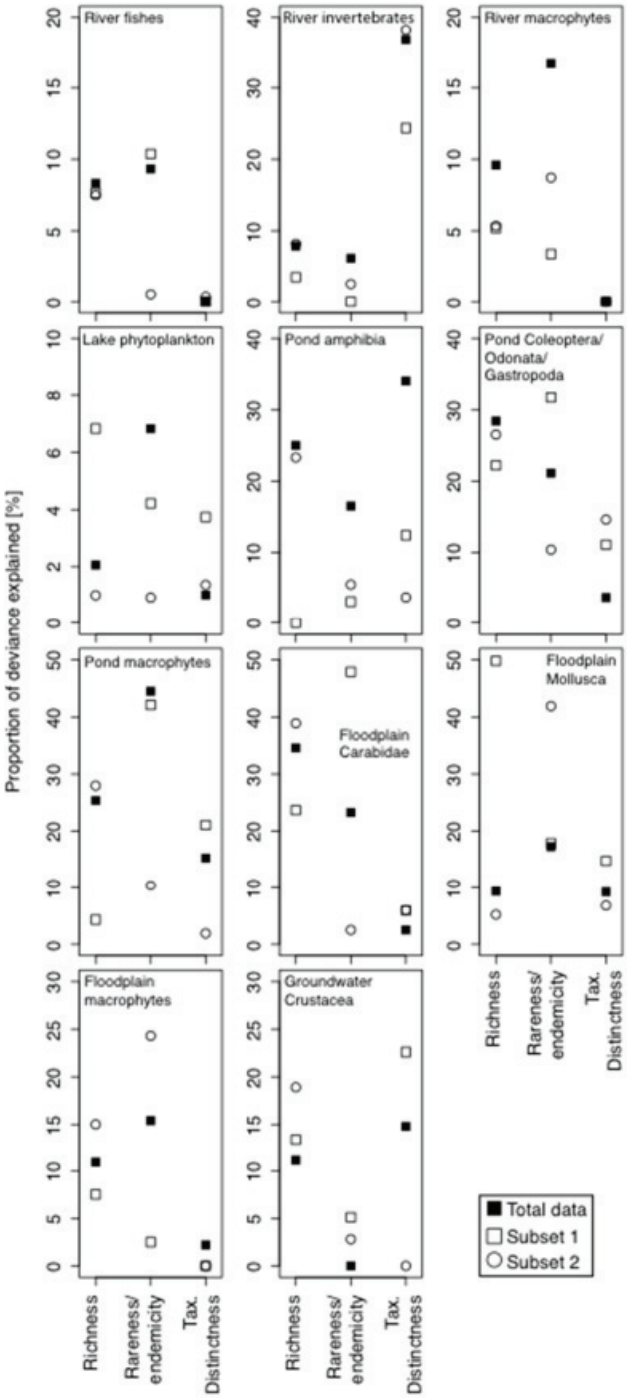


Figure 2: Proportion of deviance explained by land use and interactions with land use in the GLM models using three biodiversity metrics calculated for eleven organism groups. Each model run was repeated using the full dataset (filled symbol) and two data subsets (empty symbols). Data subsets were generated separately for each biodiversity metric and based on the split points identified by Boosted Regression Tree analysis for the strongest geo-climatic environmental descriptor variable in each model (see text for details).

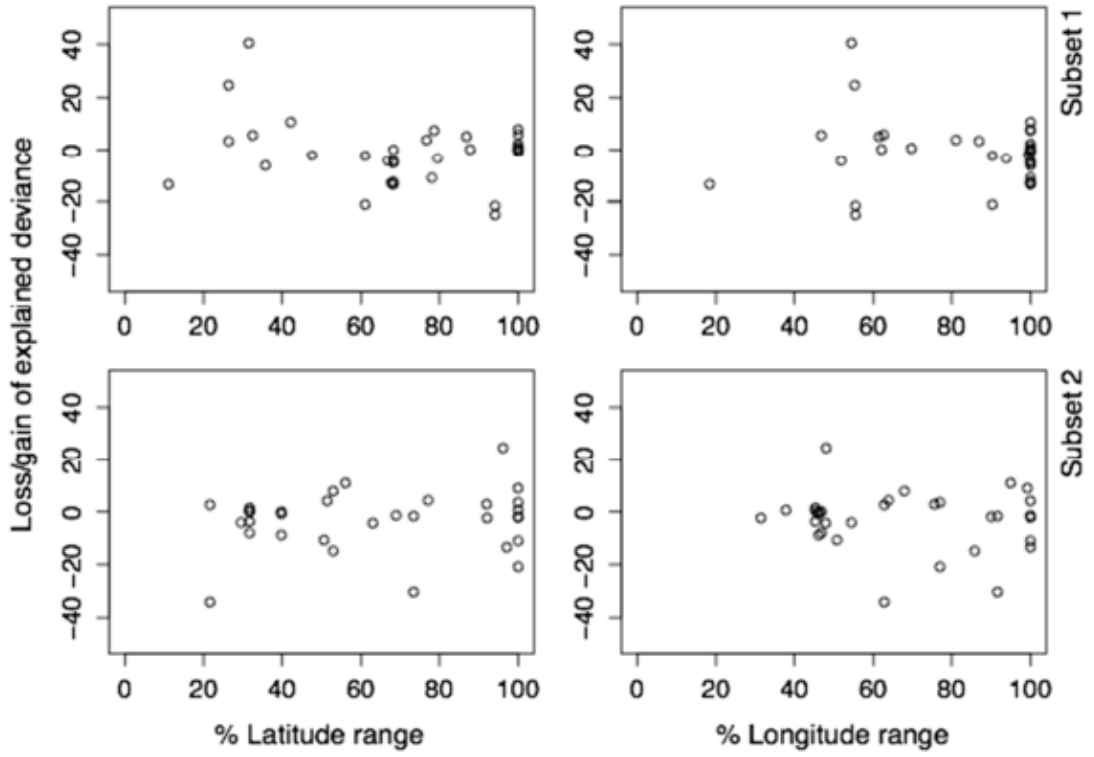


Figure 3: Changes in the proportion of deviance explained by land use (GLM models, absolute values) against percent range of latitude and longitude covered by data subsets 1 and 2 in comparison to the range of the full dataset. High percent values on the x-axis indicate a higher resemblance of latitude and longitude gradients to those of the full dataset. For the definition of subsets 1 and 2, see Table 4.

reveal inconsistent patterns, without a general increase (or decrease) of the role of human impact in the data subsets. The comparison between GLM models using the full data and the data subsets, however, reveal climatic gradients (temperature, precipitation) influence freshwater biodiversity to a greater extent than geographical gradients (latitude, longitude). Climatic and geographical gradients, although strongly linked at the continental scale, are not necessarily congruent. They reveal different patterns: while latitude and longitude represent continuous gradients from the north to the south and from the east to the west, climatic gradients are changing with altitude and other factors and hence are rather discontinuous at the European scale. The outcome of this study reveals that the role of land use increased only if the subsets were split along climatic gradients. This supports a rather discontinuous pattern of temperature and precipitation across ecoregions as compared to the geographical gradients. If we assume similar discontinuous patterns are inherent in our freshwater biodiversity data, this may explain the greater role of climatic descriptors in the full dataset too.

In summary, the data subsetting exercise highlights land use plays a stronger role in driving freshwater biodiversity in geo-climatically more homogeneous data subsets. Yet, this does not necessarily mean the subsets cover a reduced geographical extent, e.g. comparable alpine climates are found in mountainous central Europe as well as parts of flatter northern Europe. As this study is the first to address these patterns at the broad scale and across numerous freshwater ecosystem types and organism groups, future studies are required to investigate the role of spatial patterns in human land uses in respect of freshwater biodiversity responses. A key focus should be the identification of the

spatial scale best suited to detect land use and other human impacts.

The general response of freshwater biodiversity to land use

There is considerable evidence that urban (reviewed by Paul & Meyer, 2001) and agricultural (reviewed by Allan, 2004, see also Feld *et al.,* 2013) land uses adversely affect the biodiversity and integrity of lotic ecosystems. Likewise, pond macrophyte and invertebrate richness are impacted by agriculture (Declerck *et al.,* 2006; Della Bella & Laura, 2009) and pond amphibian and macrophyte richness by urbanisation (Akasaka *et al.,* 2010; Hartel *et al.,* 2010). Similar adverse effects of human land use on freshwater biodiversity are reported for lakes (Brucet at al., 2013) and obligate groundwater fauna (Malard *et al.,* 1996). For lakes, land-use change is considered the most severe driver of biodiversity change (Sala *et al.,* 2000), so that adverse effects on plankton diversity through mechanisms of 555 nutrient loading (e.g. Nielsen *et al.,* 2012) and water quality deterioration (e.g. Jeppesen *et al.,* 2000) were highly anticipated also in this study.

This general decline of biodiversity in response to agricultural and urban land uses is not fully supported by our findings. Besides the generally weak pure effect of land use, we found both positive and negative relationships between biodiversity indices and both land use groups in equal parts. This has rarely been reported from other ecosystems. Davies et al. (2007) found human population density to be positively correlated with bird richness and concluded, in agreemeent with Balmford et al. (2001), "the tendency for higher levels of human density and species richness to be favoured by similar kinds of environments […] overwhelms any negative effect of those densities on avian richness." The authors also found a positive response to high levels of agricultural land use, although whether this applies to aquatic ecosystems remains speculative.

It is uncertain whether whole-community based biodiversity metrics are suitable measures to indicate adverse land use effects. Freshwater communities are often species-rich and may dramatically change along anthropogenic impact gradients, while both species richness and evenness may remain relatively stable along the same gradient and even at its end points (Feld *et al.,* 2013). Consequently, many whole community measures of biodiversity fail to detect species turnover, which renders them poor indicators of ecosystem degradation, in particular with species-rich assemblages such as benthic macroinvertebrates. This turnover might be detected by measures of beta diversity, which was not considered here. Future studies could usefully focus on changes in species composition along environmental impact gradients using measures that quantify the spatial turnover and nestedness components of beta diversity (Baselga, 2012).

With the comparison of measures of alpha diversity, we also need to 579 take the potential methodological constraints into consideration that may hamper a comparative analysis of biodiversity patterns at the broad scale. Monitoring sampling methodology, for instance, usually aims at obtaining data for a site's quality assessment, but does not allow for sampling the whole biodiversity of a given site, in particular not if only one season is being addressed. This in particular applies to lake phytoplankton biodiversity, which is notoriously difficult to estimate (Carstensen *et al.,* 2005; Uuistalo *et al.,* 2013) due to a large number of species, many of which are usually present in very low abundance. Further, phytoplankton species (or taxon) richness is strongly linked to the sampling and counting methodology (Carstensen *et al.,* 2005) and often restricted by the use of light microscopy of preserved samples in routine monitoring schemes (Ojaveer *et al.,* 2010). Hence, there is potential for methodological inconsistency in our data, which, in part, may have caused the weak response patterns observed for lake phytoplankton, but nevertheless which resulted in the consistently weak pure effects of land use on biodiversity across organism groups and ecosystems.

Implications for monitoring freshwater biodiversity

This study posits three major conclusions, with strong implications for future research on freshwater biodiversity and its response to le anthropogenic stressors at large spatial scales:

1. The combined analysis of both geo-climatic and anthropogenic impact gradients is a prerequisite for the detection and quantification of human threats to biodiversity. Natural geo-climatic and anthropogenic factors may be collinear, jointly explain a considerable amount of (shared) variation in the response variable and interact with each other, all of which complicate the detection of biodiversity response to anthropogenic impact. As this interaction can be assumed to be inherent to any large-scale (e.g. ecoregional or continental) dataset, separating analysis of geo-climatic and anthropogenic gradients cannot account for the shared effects and interactions as this would result in erroneous interpretation of biodiversity response patterns to environmental gradients.

2. Geo-climatic descriptors form strong gradients in large-scale datasets. These gradients may mask anthropogenic gradients and thus complicate or even hinder the detection of the latter. More homogeneous datasets (with reduced gradients of natural explanatory variables) can help overcome the dominance of natural gradients and may also provide stronger models explaining more variance in the biological response variable.

3. Whole community-based biodiversity metrics, such as species richness, Shannon-Wiener Diversity, Pi.lou‘s evenenness or taxonomic distinctness show responses to anthropogenic stressor gradients, but there is sufficient evidence of contrasting response directions, with increasing as well as decreasing biodiversity values along various stressor gradients. Further, as whole-community biodiversity metrics may fail to detect the turnover in species composition, other metrics capable of accounting for species identity and turnover should be tested in addition when the biodiversity response to land use and other anthropogenic stressors in freshwater ecosystems is under consider.

Acknowledegements

This study was supported by the European Commission through the BioFresh project: FP7-ENV-2008, Contract no. 226874. We are thankful for to the following institutions and organisations for providing data for this study: Mediterranean GIG (Data manager: Caridad de Hoyos, CEDEX-CEH); Ministry of Agriculture, Natural Resources and Environment Cyprus, Water Development Department (MANRE-WDD); water agencies and Institut national de recherche en sciences et rechnologies pour environnement et agriculture (IRSTEA) and ONEMA, France; Maria Moustaka, Aristotle University of Thessaloniki, Greece; Universit. degli Studi di Sassari, Dipartimento di Scienze Botaniche, Ecologiche 628 e Geologiche (DiSBEG) Italy; Instituto da Agua, I.P. (INAG), Portugal; Ministerio de Agricultura, Alimentacion y Medio Ambiente, Centro de Estudios Hidrogr.ficos (CEDEX-CEH), Spain; Central-Baltic GIG (Data manager: Ute Mischke, IGB); National Environmental Research Institute, University of Aarhus, Denmark; Estonian Ministry of the Environment; Federal Ministry for Agriculture, Forestry, Environment and Water Management (including all Austrian Federal States); Austrian Federal Agency for Water Management; Institute of Water Ecology, Fisheries and Lake Research Austria; Office of the Provincial Government of Lower Austria; German Federal States of Bavaria, Hesse, Rhineland-Palatinate, North Rhine-Westphalia, Lower Saxony, Schleswig–Holstein, Mecklenburg-Vorpommern, Saxony-Anhalt, Saxony and

Thuringia; Latvian Environment, Geology and Meteorology Centre; EPA Lithuania; Rijkswaterstaat (RWS), the Netherlands; Institute of Environmental Protection— National Research Institute, Poland; Scottish Environment Protection Agency (SEPA) and Environment Agency for England & Wales (EA), UK; Eastern-Continental GIG (Data manager: Gabor Borics, CER); Centre for Ecological Research, Hungarian Academy of Sciences; Ministeriul Meduli s ii Paidurilor (MMP), Romania; Northern GIG (Data manager: Geoff Phillips, EA); Finnish Environment Institute (SYKE); Environment Protection Agency (EPA), Ireland; Norwegian Institute for Water Research (NIVA); Swedish University of Agricultural Sciences (SLU). The following colleagues helped with the pond data: Luk.cs Bal.zs Andr.s (Debrecen, HU), Jens Arle (Dessau, DE), Sandrine Angélibert (Allinges, FR), Regis C.r.ghino (Toulouse, FR), Valentina Della Bella (Rome, IT), Carolin Meier (Waiblingen, DE), Beat Oertli (Jussy, CH), Gwendolin Porst (Berlin, DE), Carl Sayer (London, UK), Aneta Spyra (Katowice, PL), Sonja Stendera (Solingen, DE), Nigel Willby (Stirling, UK) and Pablo Valverde (Bielefeld, DE). The following colleagues contributed to the assembly of the European groundwater crustacean data set: Maja Zagmajster (SL), Cene Fišer (SL), D. Galassi (IT), Fabio Stoch (IT) and Pierre Marmonier (FR).

References

Akasaka M., Takamura N., Mitsuhashi H. & Kadono Y. (2010) Effects of land use on aquatic macrophyte diversity and water quality of ponds. *Freshwater Biology*, **55**, 909–922.

Allan J.D. (2004). Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. *Annual Review of Ecology, Evolution, and Systematics*, **35**, 257–284.

Araujo M.B., Nogu.s-Bravo D., Diniz-Filho J.A.F., Haywood A.M., Valdes P.J. & Rahbek C.(2008) Quaternary climate changes explain diversity among reptiles and amphibians. *Ecography*, **31**, 8–15.

Balmford A., Moore J.L., Brooks T., Burgess N., Hansen L.A., Williams P. & Rahbek C. (2001) Conservation Conflicts Across Africa. *Science*, 291, 2616–2619.

Baselga A., Gomez-Rodrıguez C., Lobo J.M. (2012) Historical legacies in world amphibian diversity revealed by the turnover and nestedness components of beta diversity. *PLoS ONE*, **7**, 323–341.

Birk S., Bonne W., Borja A., Brucet S., Courrat A., Pokane S., Solimini A., van de Bund W., Zampoukas N. & Hering D. (2012) Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. *Ecological Indicators*, **18**, 31–41.

Böhmer J. (2012). Amphibienvorkommen und deren s.urebedingten Laichsch.den im Nordschwarzwald und Odenwald in 2012 im Vergleich mit den Jahren 1987-1990, 1996 und 2002 (p. 51). BioForum GmbH, Kirchheim/Teck.

Brucet S., P.dron S., Mehner T., Lauridsen T.L., Argillier C., Winfield I.J., Volta P., Emmrich M., Hesthagen T., Holmgren K., Benejam L., Kelly F., Krause T., Palm A., Rask M. & Jeppesen E. (2013) Fish diversity in European lakes: geographical factors dominate over anthropogenic pressures. *Freshwater Biology*, **58**, 1779–1793.

Carstensen J., Frohn L.M., Hasager C.B. & Gustafsson B.G. (2005) Summer algal blooms in a coastal ecosystem: the role of atmospheric deposition versus entrainment fluxes. *Estuarine, Coastal and Shelf Science*, **62**, 595–608.

Clapcott J.E., Collier K.J., Death R.G., Goodwin E.O., Harding J.S., Kelly D., Leathwick J.R., Young R.G. (2012) Quantifying relationships between land-use gradients and structural and functional indicators of stream ecological integrity. *Freshwater Biology*, 57, 74–90.

Clarke K. & Warwick R. (1998) a taxonomic distinctness index and its statistical properties. *Journal of Applied Ecology*, **35**, 523–531.

Clarke K. & Warwick R. (1999) The taxonomic distinctness measure 688 of biodiversity: weighting of step lengths between hierarchical levels. *Marine Ecology Progress Series*, **184**, 21–29.

Crisp D.T., Laffan S., Linder H.P. & Monro A. (2001) Endemism in the Australian flora. *Journal of Biogeography*, **28**, 183–198.

Cutler D. R., Edwards T.C., Beard K.H., Cutler A., Hess K.T., Gibson J. & Lawler J.J. (2007) Random forests for classification in ecology. *Ecology*, **88**, 2783–2792.

Dahm V., Hering D., Nemitz D., Graf W., Schmidt-Kloiber A., Leitner P., Melcher A., Feld C.K. (2013) Effects of physico-chemistry, land use and hydromorphology on three riverine organism groups: a comparative analysis with monitoring data from Germany and Austria. *Hydrobiologia*, **704**, 389–415.

Davies R.G., Orme C.D.L., Olson V., Thomas G.H., Ross S.G., Ding T.-S., et al. (2006) Human impacts and the global distribution of extinction risk. *Proceedings Biological Sciences / The Royal Society*, **273**, 2127–2133.

Davies R.G., Orme C.D.L., Storch D., Olson V.A., Thomas G.H., Ross S.G., Ding T.S., Rasmussen P.C., Bennett P.M., Owens I.P.F., Blackburn T.M. & Gaston K.J. (2007) Topography, energy and the global distribution of bird species richness. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 1189–1197.

De Meester L., Declerck S., Stoks R., Louette G., Van De Meutter F., De Bie T., Michels E. & Brendonck L. (2005). Ponds and pools as model systems in conservation biology, ecology and evolutionary biology. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **15**, 715–725.

Declerck S., De Bie T., Ercken D., Hampel H., Schrijvers S., Van Wichelen J., Gillard V., Mandiki R., Losson B., Bauwens D., Keijers S., Vyverman W., Goddeeris B., De Meester L., Brendonck L. & Martens K. (2006). Ecological characteristics of small farmland ponds: associations with land use practices at multiple spatial scales. *Biological Conservation*, **131**, 523–532.

Deharveng L., Stoch F., Gibert J., Bedos A., Galassi D., Zagmajster M., Brancelj A., Camacho A., Fiers F., Martin P., Giani N., Magniez G. & Marmonier P. (2009) Groundwater biodiversity in Europe. *Freshwater Biology*, **54**, 709–726.

Della Bella V. & Mancini L. (2009). Freshwater diatom and macroinvertebrate diversity of coastal permanent ponds along a gradient of human impact in a Mediterranean ecoregion. *Hydrobiologia*, **634**, 25–41.

Dudgeon D., Arthington A.H., Gessner M.O., Kawabata Z.-I., Knowler 721 D.J., Lévêque C., Naiman R.J., Prieur-Richard A.-H., Soto D., Stiassny M.L.J. & Sullivan C.A. (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews*, **81**, 163.

Elith J., Leathwick J. R. & Hastie T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.

Ellenberg H., Weber H. E., Düll R., Wirth V., Werner W. & Pauli.en D. (1992). Indicator values of plants in Central Europe (Vol. 18, p. 258). Erich Goltze Verlag, Göttingen.

Evans K.L., Warren P.H. & Gaston K.J. (2005) Species–energy relationships at the macroecological scale: a review of the mechanisms. *Biological Reviews,* **80**, 1–25.

Feld, C.K. (2013) Response of three lotic assemblages to riparian and catchment-scale land use: implications for designing catchment monitoring programmes. *Freshwater Biology*, **58**, 715–729.

Feld C.K., Birk S., Bradley D.C., Hering D., Kail J., Marzin A., Melcher A., Nemitz D., Pedersen M.L., Pletterbauer F., Pont D., Verdonschot P.F.M. & Friberg N. (2011) Chapter- From Natural to Degraded Rivers and Back Again: a Test of Restoration Ecology Theory and Practice. *Advances in Ecological Research*, **44**, 119–209.

Feld C.K., de Bello F. & Dol.dec S. (2013) Biodiversity of traits and species both show weak responses to hydromorphological alteration in lowland river macroinvertebrates. *Freshwater Biology*, **59**, 233–248.

Field R., Hawkins B. a., Cornell H. V., Currie D.J., Diniz-Filho J.A.F., Gu.gan J.-F., et al. (2009) Spatial species-richness gradients across scales: a meta-analysis. *Journal of Biogeography*, **36**, 132–147.

Gallardo B., Gasc.n S., Quintana X., Com.n F.A. (2011) How to choose a biodiversity indicator – Redundancy and complementarity of biodiversity metrics in a freshwater ecosystem. *Ecological Indicators*, **11**, 1177–1184.

Gouveia S.F., Hortal J., Cassemiro F. a. S., Rangel T.F. & Diniz-Filho J.A.F. (2013) Nonstationary effects of productivity, seasonality, and historical climate changes on global amphibian diversity. *Ecography*, **36**, 104–113.

Guégan J.-F., Lek S. & Oberdorff T. (1998) Energy availability and habitat heterogeneity predict global riverine fish diversity. *Nature*, **391**, 382–384.

Harding J.S., Benfield E.F., Bolstad P.V., Helfman G.S. & Jones E.B.D. III. (1998) Stream biodiversity: The ghost of land use past. *Proceedings of the National Academy of Science USA*, **95**, 14843–14847.

Hartel T., Schweiger O., Öllerer K., Cogălniceanu D. & Arntzen J. W. 755 (2010). Amphibian distribution in a traditionally managed rural landscape of Eastern Europe: Probing the effect of landscape composition. *Biological Conservation*, **143**, 1118–1124.

Hawkins B.A., Field R., Cornell H. V., Currie D.J., Gu.gan J.-F., Kaufman D.M., Kerr J.T., Mittelbach G.G., Oberdorff T., O'Brien E.M., Porter E.E. & Turner J.R.G (2003) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105–3117.

Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G. & Jarvis A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hijmans R.J., Phillips S., Leathwick J. & Elith J. (2013) dismo: Species Distribution Modeling. R Package Version 0.8-11. Available at: http://CRAN.R766 project.org/package=dismo.

Homes V., Hering D. & Reich M. (1999). The distribution and macrofauna of ponds in stretches of an alpine floodplain differently impacted by hydrological engineering. *Regulated Rivers: Research & Management*, **417**, 405–417.

Hortal J., Diniz-Filho J.A.F., Bini L.M., Rodr.guez M..., Baselga A., Nogu.s-Bravo D., et al. (2011) Ice age climate, evolutionary constraints and diversity patterns of European dung beetles. *Ecology Letters*, **14**, 741–748.

Illies J. (1978) Limnofauna Europaea. 2. Auflage. Gustav Fischer Verlag, New York, Stuttgart. 532 pp.

Jeppesen E., Meerhoff M., Holmgren K., Gonzalez-Bergonzoni I., Teixeira-de Mello F., Declerck A.A.J. et al. (2010) Impacts of climate warming on lake fish community structure and potential effects on ecosystem function. *Hydrobiologia*, **646**, 73–90.

Jeppesen E., Mehner T., Winfield I.J., Kangur K., Sarvala J., Gerdeaux D., et al. (2012) Impacts of climate warming on the long-term dynamics of key fish species in 24 European lakes. *Hydrobiologia*, **694**, 1–3.

Jeppesen E., Peder Jensen J., Sondergaard M., Lauridsen T. & Landkildehus F. (2000). Trophic structure, species richness and biodiversity in Danish lakes: changes along a phosphorus gradient. *Freshwater Biology*, **45**, 201–218.

Knight L. (2012) Hypogean Crustacea Recording Scheme. Data available at: http://hcrs.freshwaterlife.org.

Legendre P. & Legendre L. (1998) *Numerical Ecology*, 2nd edition. Elsevier, Amsterdam.

Leprieur F., Tedesco P., Hugueny B., Beauchard O., Dürr H.H., Brosse S., et al. (2011) Partitioning global patterns of freshwater fish beta diversity reveals contrasting signatures of past climate changes. Ecology Letters, 14, 325–334.

Linder H.P. (2001) Plant diversity and endemism in sub-Saharan tropical Africa. *Journal of Biogeography*, **28**, 169–182.

Loh J. & Wackernagel M. (ed.) (2004) Living Planet Report 2004. World-Wide Fund for Nature International (WWF), Global Footprint Network, UNEP World Conservation Monitoring Centre, Gland, Switzerland.

MacArthur R.H. & Wilson E.O (1967) The theory of island biogeography. Princeton University Press, Princeton, New Jersey, USA.

Malard F., Mathieu J., Reygrobellet J.-L. & Lafont M. (1996) Biomonitoring groundwater contamination: application to a karst area in Southern France. *Aquatic Sciences*, **58**, 158–799 187.

MEA (Millennium Ecosystem Assessment) (2005). Ecosystems and Human Well-being: Synthesis. Island Press, Washington, DC.

Mittelbach G.G., Schemske D.W., Cornell H.V, Allen A.P., Brown J.M., Bush M.B., et al. (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, **10**, 315–331.

Moe S.J., Schmidt-Kloiber A., Dudley B.J. & Hering D. (2013). The WISER way of organising ecological data from European rivers, lakes, transitional and coastal waters. *Hydrobiologia*, **704**, 11–28.

Myers N., Mittermeier R.A., Mittermeier C.G., da Fonseca G.A.B. & Kent J. (2007) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Nagorskaya L., Moroz M., Laeno T., Veznovetz V., Moller Pillot H., Dijkstra K.D.B. & Reemer M. (2002). Macrofauna in floodplain pools and dead branches of the Pripyat river, Minsk. Institute of Zoology, National Academy of Science Belarus, p. 158.

Nielsen A., Trolle D., S.ndergaard M., Lauridsen T.L., Bjerring R., Olesen J.E., Jeppesen E. (2012)

Watershed land use effects on lake water quality in Denmark. *Ecological Applications*, **22**, 1187–1200.

Oberdorff T., Tedesco P., Hugueny B., Leprieur F., Beauchard O., Brosse S., et al. (2011) Global and Regional Patterns in Riverine Fish Species Richness: A Review. *International Journal of Ecology*, **2011**, 1–12.

Oertli B., Auderset Joye D., Castella E., Juge R., Lehmann A. & Lachavanne 819 J.-B. (2005) PLOCH: a standardized method for sampling and assessing the biodiversity in ponds. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **15**, 665–679.

Ojaveer H., Jaanus A., MacKenzie B.R., Martin G., Olenin S., Radziejewska T., Telesh I., Zettler M.L. & Zaiko A. (2010) Status of Biodiversity in the Baltic Sea. *PLoS ONE*, **5**, e12467.

Paul M.J. & Meyer J.L. (2001) Streams in the urban landscape. *Annual Review of Ecology and Systematics*, **32**, 333–365.

Phillips G., Lyche-Solheim, A., Skjelbred B., Mischke U., Drakare S., Free G., J.rvinen M., de Hoyos C., Morabito G., Poikane S. & Carvalho L. (2012) A phytoplankton trophic index to assess the status of lakes for the Water Framework Directive. *Hydrobiologia*, **704**, 75–95.

R Development Core Team (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org. (Version: 2.15.3)

Reyjol Y., Hugueny B., Pont D., Bianco P.G., Beier U., Caiola N., Casals F. Cowx I.G., Economou A., Ferreira M. T., Haidvogl G., Noble R., de Sostoa A., Vigneron T. & Virbickas T. (2007). Patterns in species richness and endemism of European freshwater fish. *Global Ecology and Biogeography*, **16**, 65–75.

Ridgeway G. (2013) gbm: Generalized Boosted Regression Models R Package Version 2.0-8. Available at: http://CRAN.Rproject. org/package=gbm.

Ruffo S. & Stoch F. (2006) Checklist and distribution of the Italian fauna. 10,000 terrestrial and inland waters species. Memorie del Museo Civico di Storia Naturale di Verona, 2. serie. *Sezione Scienze della Vita*, **17**, 1–307. (data on CD-ROM)

Sala O.E., Chapin F.S. III, Armesto J.J., Berlow E., Bloomfield J., Dirzo R., Huber-Sanwald E., Huenneke L.F., Jackson R.B., Kinzig A., Leemans R., Lodge D.M., Mooney H.A., Oesterheld M., Poff N.L., Sykes M.T., Walker B.H., Walker M. & Wall D.H. (2000) Global Biodiversity Scenarios for the Year 2100. *Science*, **287**, 1770–1774.

Sastre P., Roca P., Lobo J.M. & EDIT co-workers (2009) A Geoplatform for the accessibility to environmental cartography. *Journal of Biogeography*, **36**, 568.

Sayer C.D., Davidson T. & Jones J. I. (2010). Seasonal dynamics of macrophytes and phytoplankton in shallow lakes: a eutrophication-driven pathway from plants to plankton? *Freshwater Biology*, **55**, 500–513.

Scheuerell M.D. & Schindler D.E. (2004) Changes in the spatial distribution 852 of fishes in lakes along a residential development gradient. *Ecosystems*, **7**, 98 –106.

Schindler D.W. (2006) Recent advances in the understanding and management of eutrophication. *Limnology and Oceanography*, **51**, 356 –363.

Sobkowiak S. (2003) Vergleichende Untersuchungen zur Makroinvertebratenfauna von Auengew.ssern – unter besonderer Berücksichtigung der Kinzig-Aue (Hessen). Diploma thesis,

University of Duisburg-Essen, Germany.

Stendera S., Adrian R., Bonada N., Ca.edo-Argüelles M., Hugueny B., Januschke K., Pletterbauer F., & Hering D. (2012) Drivers and stressors of freshwater biodiversity patterns across different ecosystems and scales: a review. *Hydrobiologia*, **696**, 1–28.

Tisseuil C., Cornu J.-F., Beauchard O., Brosse S., Darwall W., Holland R., Hugueny B., Tedesco P.A. & Oberdorff T. (2013) Global diversity patterns and cross-taxa convergence in freshwater systems. *Journal of Animal Ecology*, **82**, 365–376.

Uusitalo L., Fleming-Lehtinen V., H.llfors H., Jaanus A., H.llfors S. & London L. (2013) A novel approach for estimating phytoplankton biodiversity. *ICES Journal of Marine Science*. doi:10.1093/icesjms/fss198

Venables W.N. & Ripley B.D. (2002) Modern Applied Statistics with S (MASS). 4th Ed.Springer, New York. Available: http://cran.r-project.org/web/packages/MASS/index.html. (Accessed on 1 April 2014).

Vörösmarty C.J., McIntyre P.B., Gessner M.O., Dudgeon D., Prusevich A., Green P., Glidden S., Bunn S.E., Sullivan C.A., Reidy Liermann C. & Davies P.M. (2010) Global threats to human water security and river biodiversity. *Nature*, **467**, 555–561.

Warton D.I. & Hui F.K.C. (2011) The arcsine is asinine: the analysis of proportions in ecology. *Ecology*, **92**, 3–10.

Wetzel R. (2001) Limnology—Lake and River Ecosystems. Academic Press, New York.

Wright D.H. (1983). Species-energy theory: an extension of species-area theory. *Oikos*, **41**, 496–506.

Zomer R.J, Trabucco A,, Bossio D.A., van Straaten, O. & Verchot L.V. (2008) Climate Change Mitigation: A Spatial Analysis of Global Land Suitability for Clean Development Mechanism Afforestation and Reforestation. *Agriculture, Ecosystems and Environment*, **126**, 67–80.

Zuur A.F., Ieno E.N. & Smith G.M. (2007) Analysing ecological data. Springer, New York, 672 pp.

Supplementary Material

Table S1: Non-collinear geo-climatic variables used as descriptors in the Boosted Regression

Trees and Generalised Linear Modelling.

Descriptors	Lakes	Rivers	Ponds	Wetlands/ floodplains	Groundwater
Longitude (°E)	x	x	x	x	x
Latitude (°N)		x			
Altitude (m a.s.l.)	x		x	x	x
Mean annual temperature (°C)	x	x	x	x	x
Annual precipitation (mm)	x	x	x	x	x
Catchment size (km^2)		x			
Actual evapo-transpiration (mm)					x
Potential evapo-transpiration (mm)			x		
Surface area (km^2)	x				

Titre Approches macro-écologique et phylogéographique pour démêler facteurs et processus responsables des patrons de biodiversité aquatique souterraine en Europe.

Résumé Un ensemble de disciplines tente de comprendre les causes de la distribution de la biodiversité à la surface de la terre. Cette thèse, à l'interface entre macro-écologie et phylogéographie, démêle le rôle relatif des différents facteurs environnementaux et des processus contrôlant la diversité des crustacés aquatiques souterrains en Europe. L'utilisation d'un modèle biologique souterrain permet d'écarter l'effet de la saisonnalité thermique, omniprésente dans les milieux de surface. L'action de multiples facteurs – plus particulièrement la disponibilité des ressources trophiques et l'hétérogénéité environnementale – et les variations régionales de leur importance relative fournissent l'explication la plus parcimonieuse des patrons de richesse. Ce résultat s'oppose au paradigme du rôle prépondérant du processus d'extinction causé par les fortes oscillations climatiques du Pléistocène en Europe du nord. Toutefois, ces oscillations ont très probablement sélectionné des organismes mobiles qui participent à l'augmentation de la taille moyenne des aires de répartition des espèces avec la latitude. La reconstruction de la dynamique des aires de distribution montre que la dispersion est un processus très hétérogène entre et au sein des espèces. Elle interviendrait lors de courtes fenêtres temporelles entre lesquelles l'adaptation locale tendrait au contraire, à contrecarrer les capacités de dispersion. Enfin, ce travail propose des pistes de réflexion afin d'expliquer plus précisément, à partir de données moléculaires supplémentaires et d'outils génomiques, les variations géographiques des taux de diversification et de substitution à l'échelle continentale.

Mots clés Patrons de biodiversité, macro-écologie, phylogéographie, crustacés, eau souterraine, carte d'habitat souterrain, reconstruction d'aire ancestrale, dispersion, adaptation locale, multi-causalité, non-stationnarité, richesse spécifique, taille des aires de distribution, beta diversité, niche réalisée/fondamentale, diversité cachée, *Proasellus*, Europe.

Title Methods in macroecology and phylogeography for disentangling factors and processes shaping groundwater biodiversity patterns in Europe.

Abstract A set of disciplines attempt to understand causes of biodiversity patterns on the earth. This thesis, at the frontier between macroecology and phylogeography, disentangles the relative influence of environmental factors and processes shaping groundwater crustacean diversity in Europe. Groundwater habitats offer useful case studies for avoiding the effect of thermal seasonality, which is pervasive in surface ecosystems. The influence of multiple factors – especially productive energy and spatial heterogeneity – and regional variation in their relative importance provide the most parsimonious explanation of species richness patterns. This result undermines the prominent role attributed to the disproportionate extinction of species in northern European regions with high historical climate oscillations. However, these oscillations have probably selected vagile species which contribute to the increase in median range size of species with latitude. Reconstructing range dynamics shows that dispersal is a heterogeneous process within and among species. It may occur during short time windows between which local adaptation favors specialization. Finally, I suggest several research avenues using molecular data and genomic tools for understanding geographical variation in diversification and substitution rates at continental scale.

Keywords: Biodiversity patterns, macroecology, phylogeography, crustaceans, groundwater, groundwater habitat map, ancestral range reconstruction, dispersal, local adaptation, multi-causality, non-stationnarity, richness, range size, beta diversity, realized/fundamental niches, cryptic diversity, *Proasellus*, Europe.